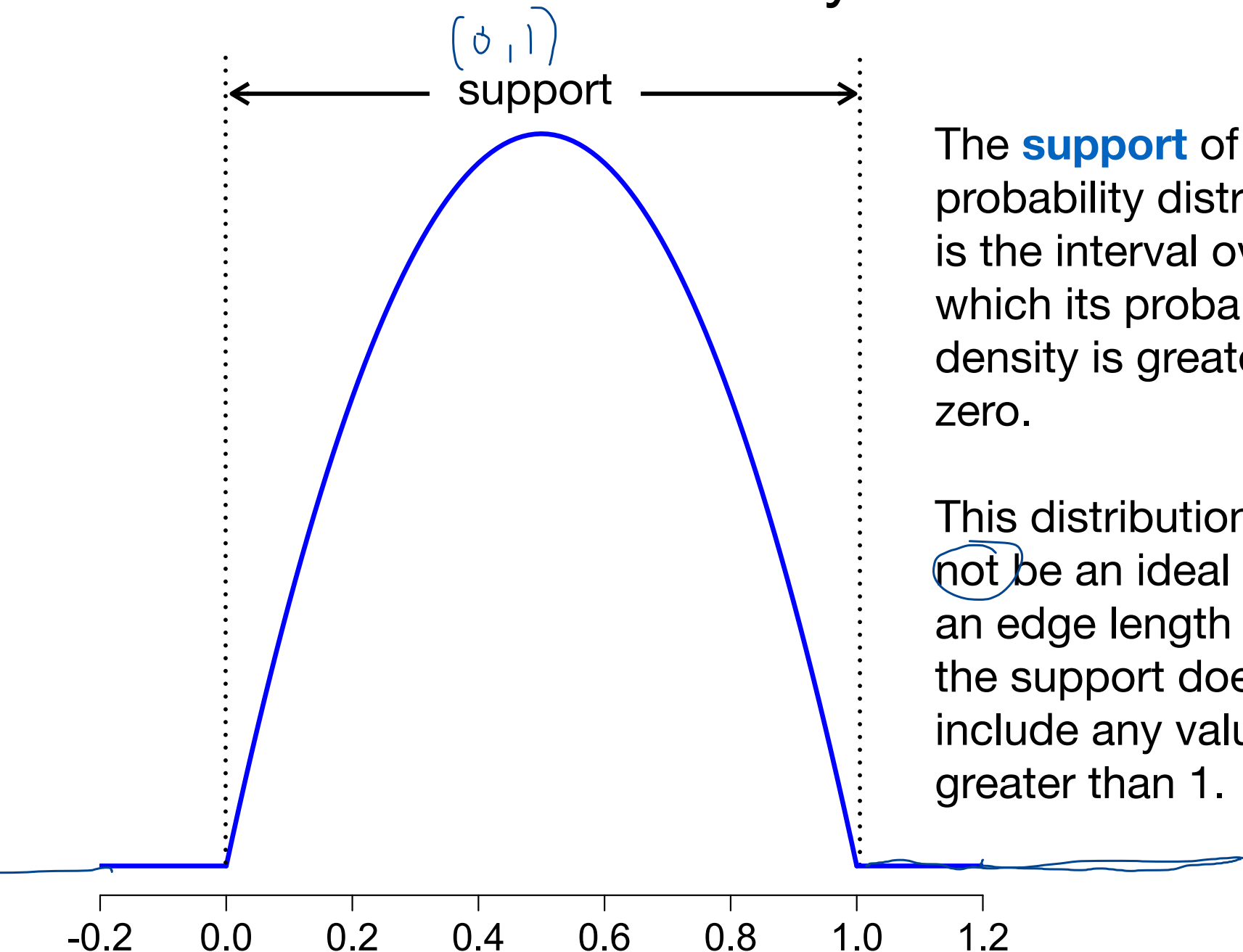


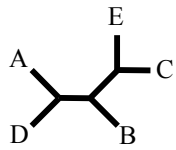
Interval in which density > 0 is the **support**



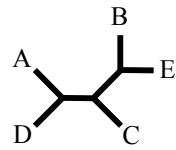
Prior Distributions

- **Discrete uniform** for topologies
- **Beta** for proportions (support $(0, 1)$) *p invar*
 - Uniform is common special case
- **Gamma** for edge lengths, rate ratios, and other parameters with support $[0, \infty)$ \leftarrow
 - Exponential is common special case
- Lognormal is alternative to Gamma
- Dirichlet for state frequencies and GTR relative rates

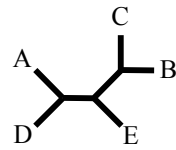
For **topologies**, a **discrete Uniform distribution** is common



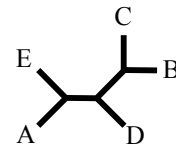
$$\frac{1}{15}$$



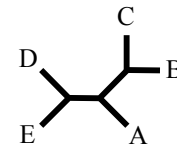
$$\frac{1}{15}$$



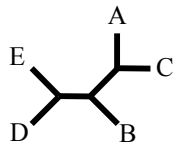
$$\frac{1}{15}$$



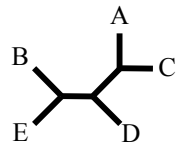
$$\frac{1}{15}$$



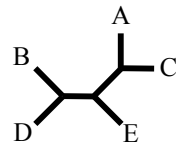
$$\frac{1}{15}$$



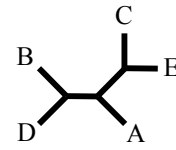
$$\frac{1}{15}$$



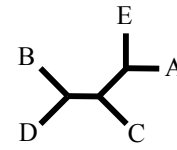
$$\frac{1}{15}$$



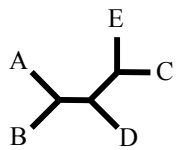
$$\frac{1}{15}$$



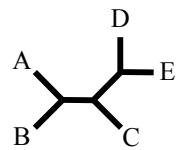
$$\frac{1}{15}$$



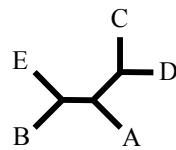
$$\frac{1}{15}$$



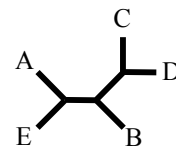
$$\frac{1}{15}$$



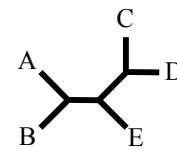
$$\frac{1}{15}$$



$$\frac{1}{15}$$

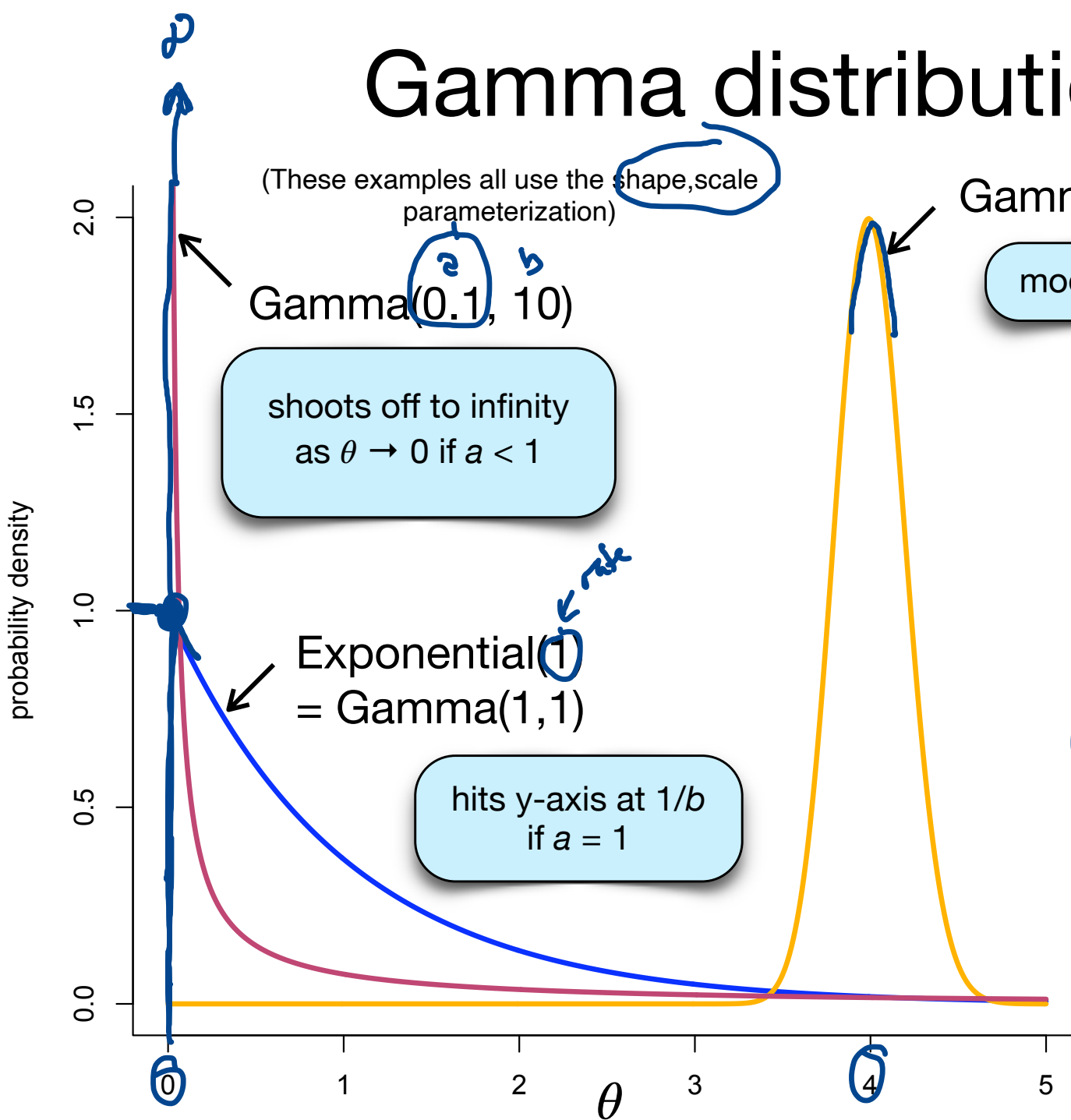


$$\frac{1}{15}$$



$$\frac{1}{15}$$

Gamma distribution



(These examples all use the shape, scale parameterization)

Gamma(a, b)

shoots off to infinity as $\theta \rightarrow 0$ if $a < 1$

Exponential(1) = Gamma(1, 1)

hits y-axis at $1/b$ if $a = 1$

Gamma(2, 0.01)

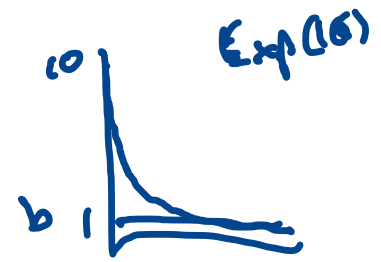
mode > 0 if $a > 1$

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(5) = Gamma(1, 1/5)



Log-normal distribution

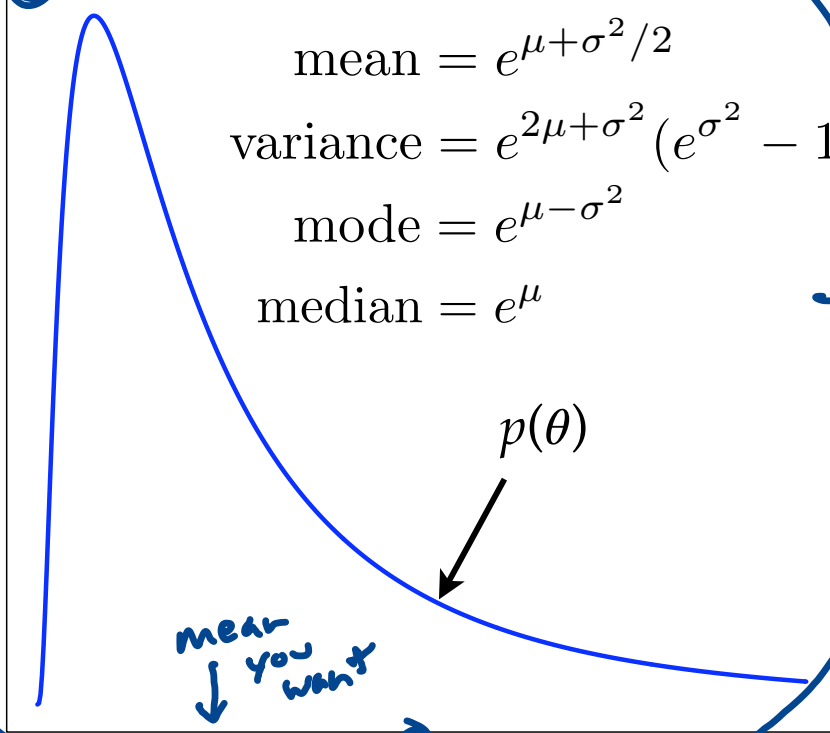
If θ is **log-normal** with parameters μ and σ .

$$\text{mean} = e^{\mu + \sigma^2/2}$$

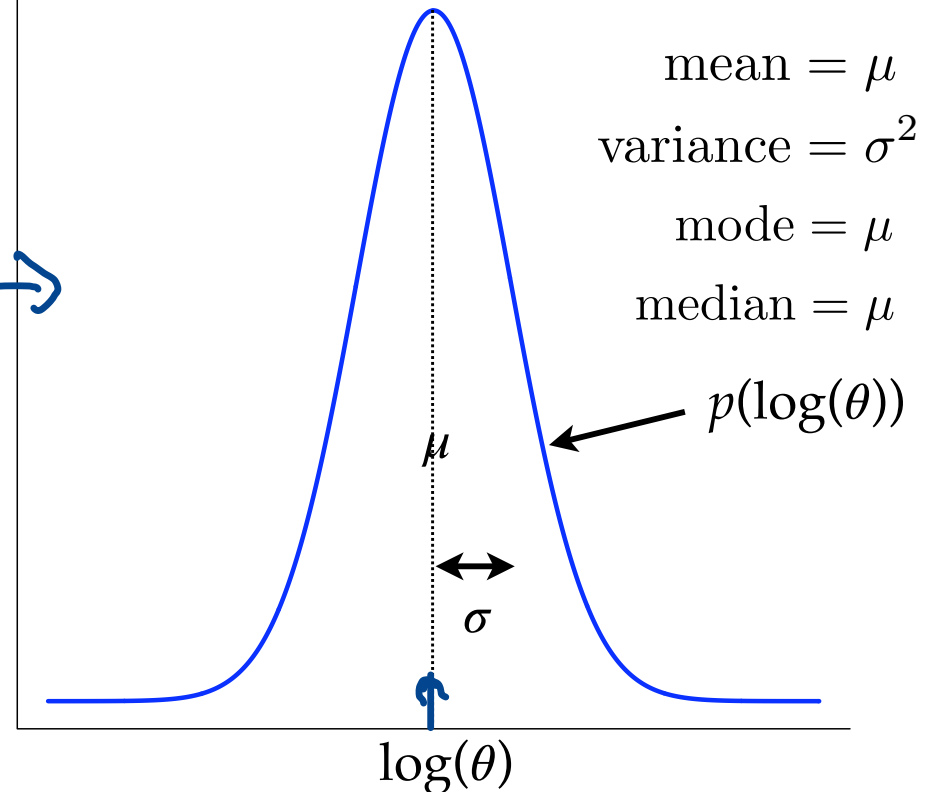
$$\text{variance} = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

$$\text{mode} = e^{\mu - \sigma^2}$$

$$\text{median} = e^{\mu}$$



...then **log(theta)** 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 $\log(\theta)$!

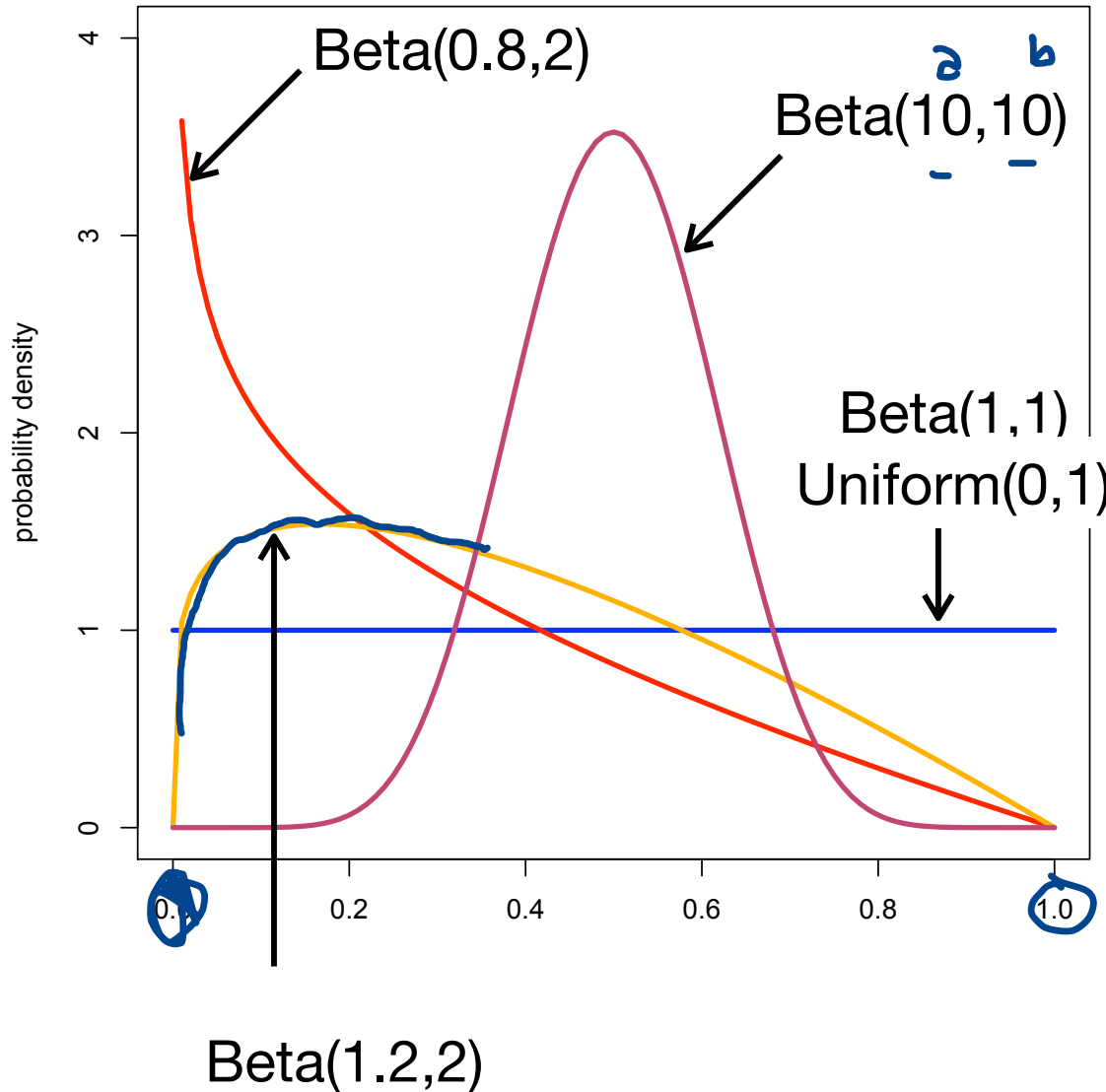
To choose μ and σ to yield a particular mean (m) and variance (v) for θ , use these formulas:

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

Handwritten notes: 'variance you want' with an arrow pointing to v, and 'mean you want' with an arrow pointing to m.

Beta distribution

1, 100
 mean = $\frac{1}{101}$



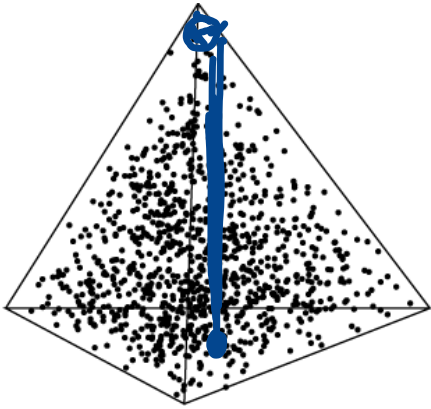
mean

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

variance

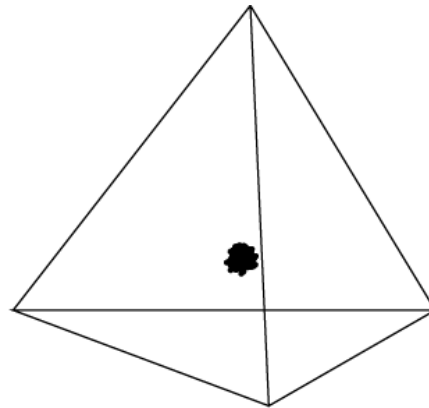
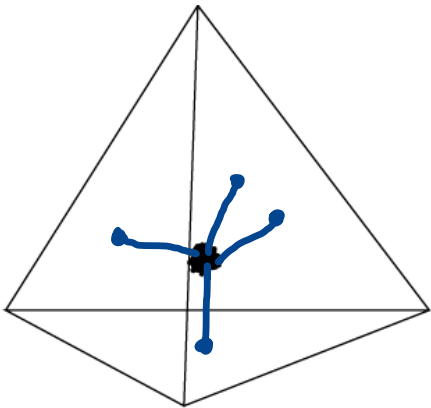
$$\frac{(\text{mean})(1 - \text{mean})}{a + b + 1}$$

Dirichlet distribution for nucleotide frequencies



Flat prior:

$$a = b = c = d = 1$$



Informative prior:

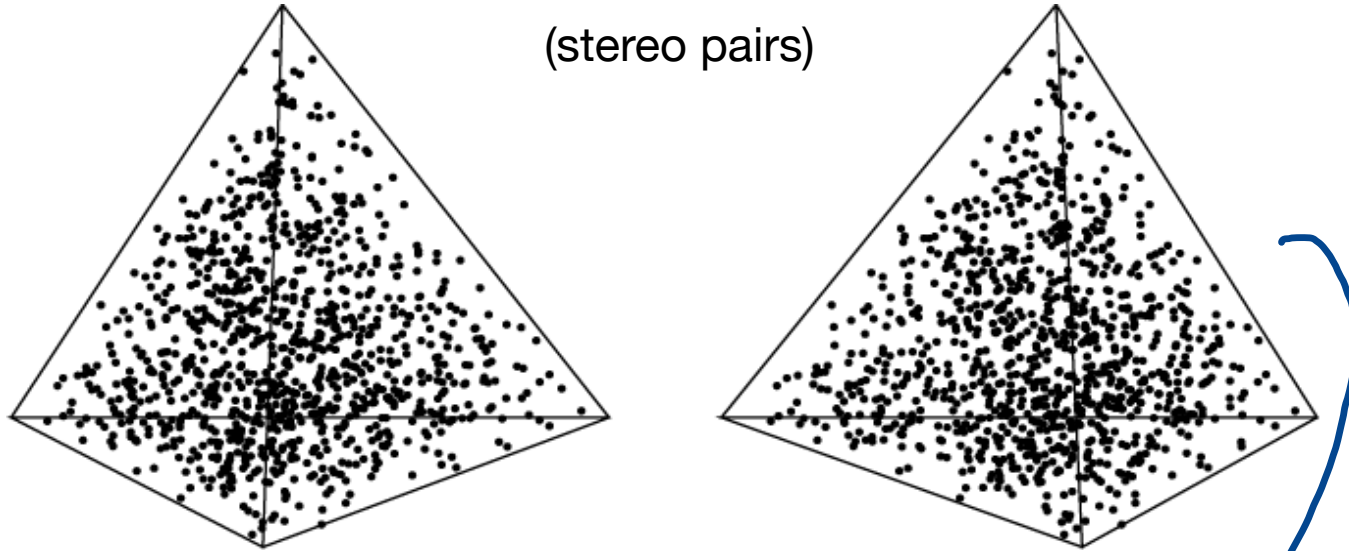
$$a = b = c = d = 300$$

(stereo pairs)

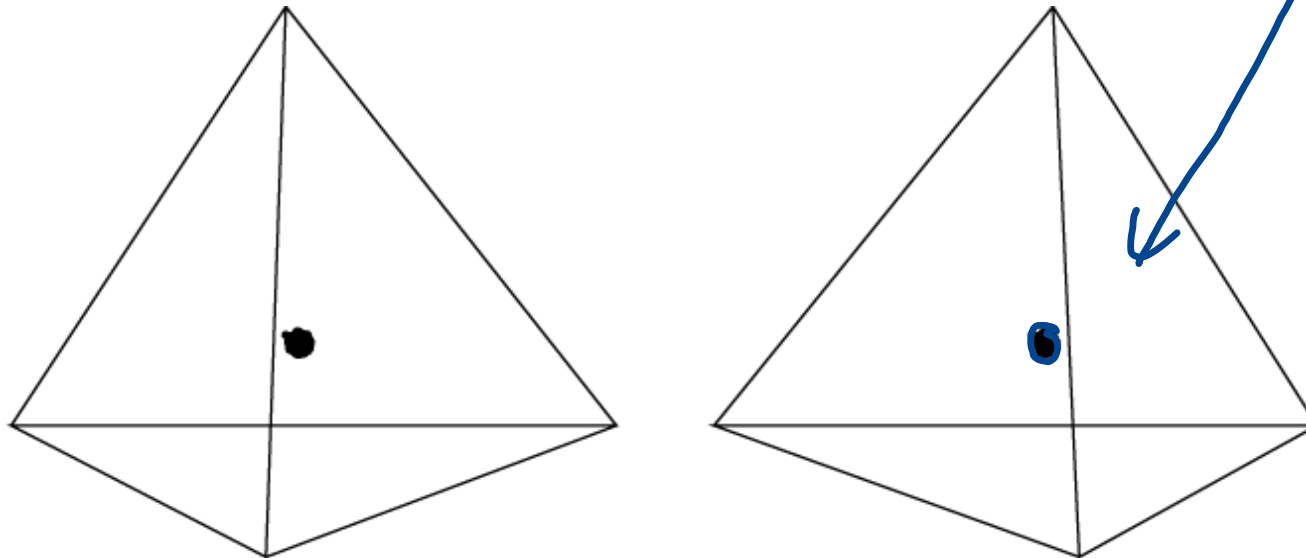
Is there information in data about nucleotide frequencies?

(stereo pairs)

Prior



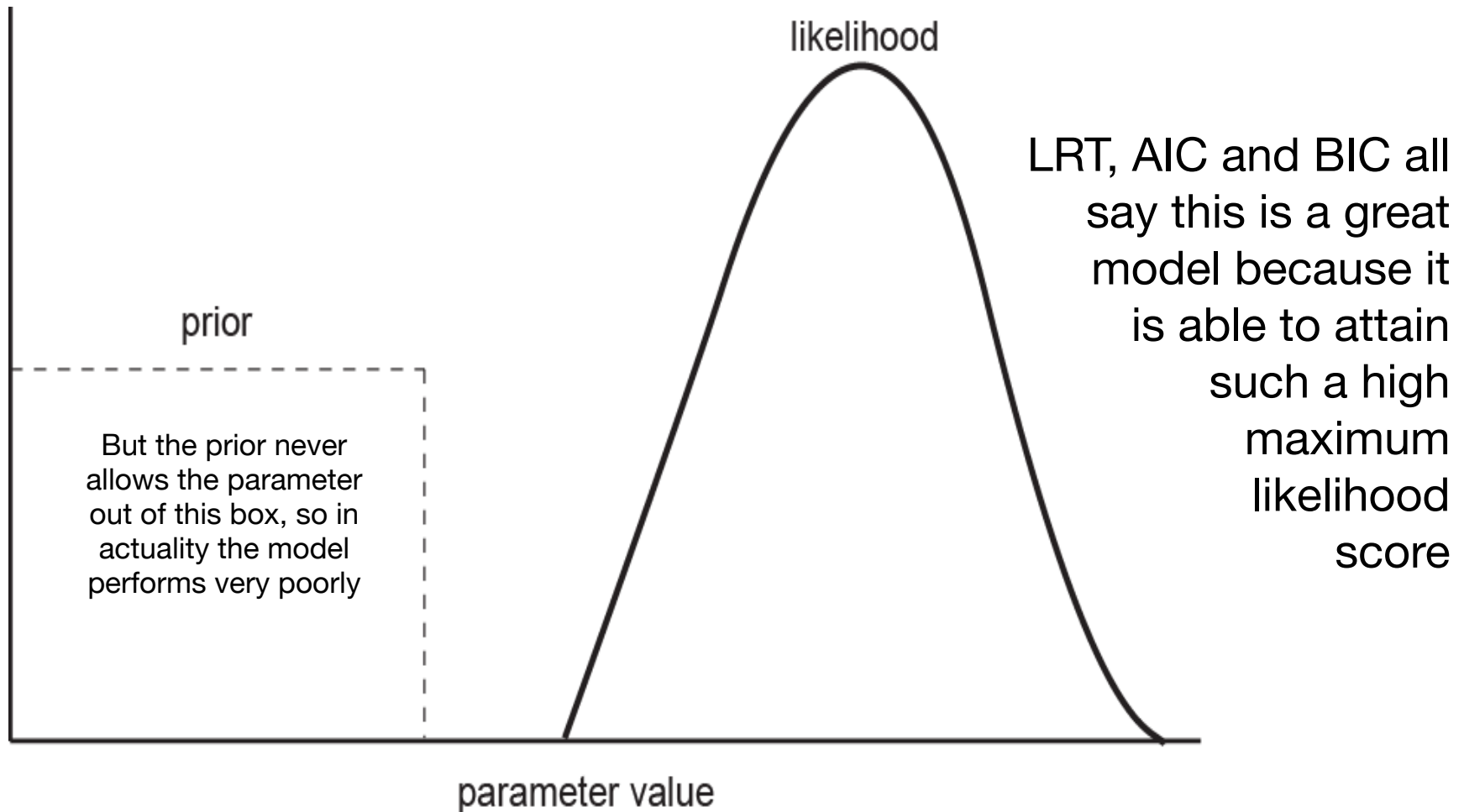
Posterior



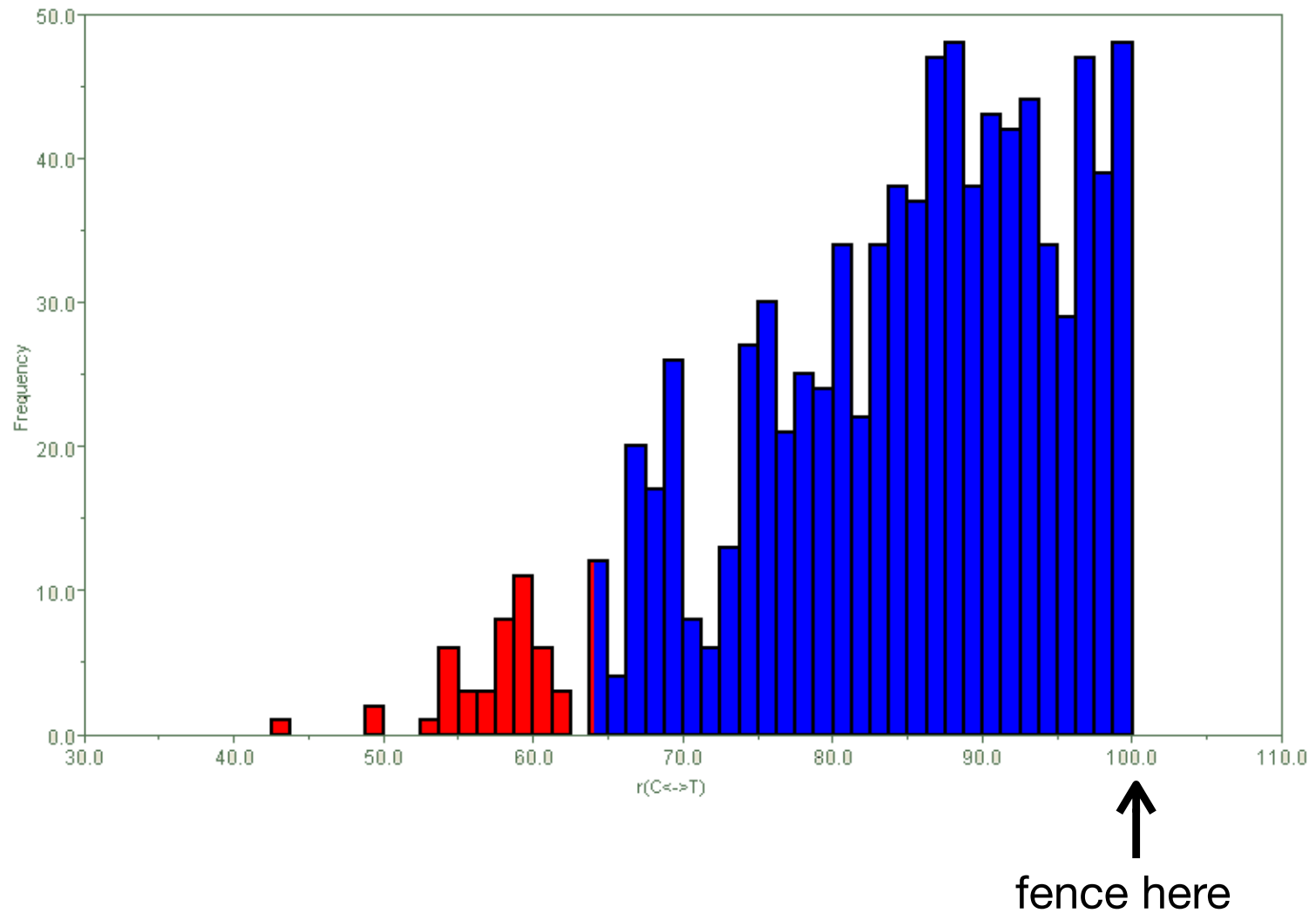
1688 aligned nucleotide sites for 34 taxa

Mending (prior) fences

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

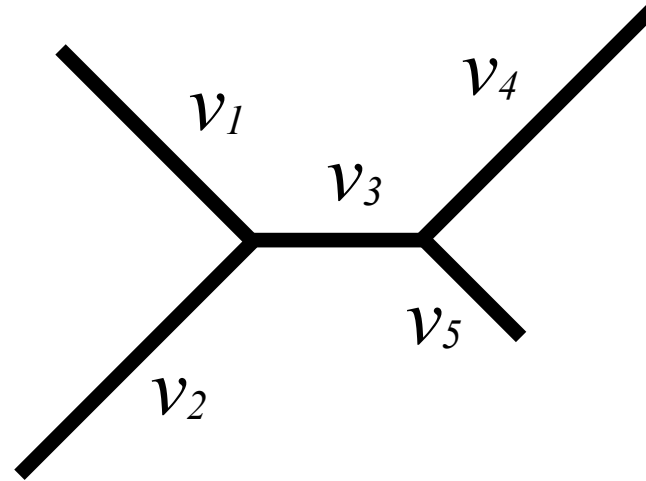


Marginal posterior distribution of r_{CT}



Beware induced priors

Induced tree length prior



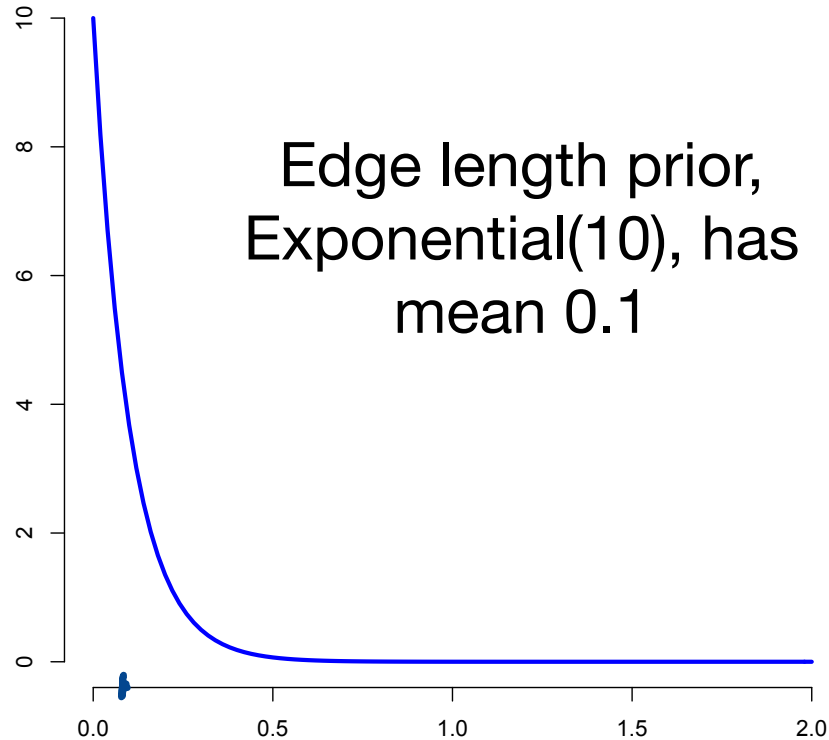
$$T = v_1 + v_2 + v_3 + v_4 + v_5$$

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

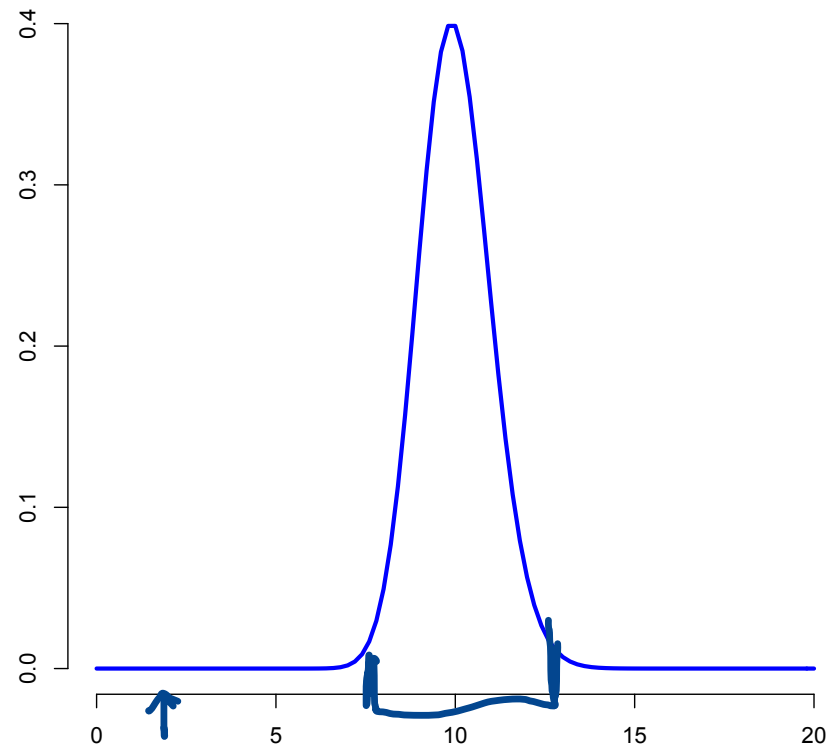
$$v_i \sim \text{Exponential}(\lambda) \longleftarrow \text{Mean} = 1/\lambda \quad \frac{1}{10} = .1$$

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

Edge length vs. tree length

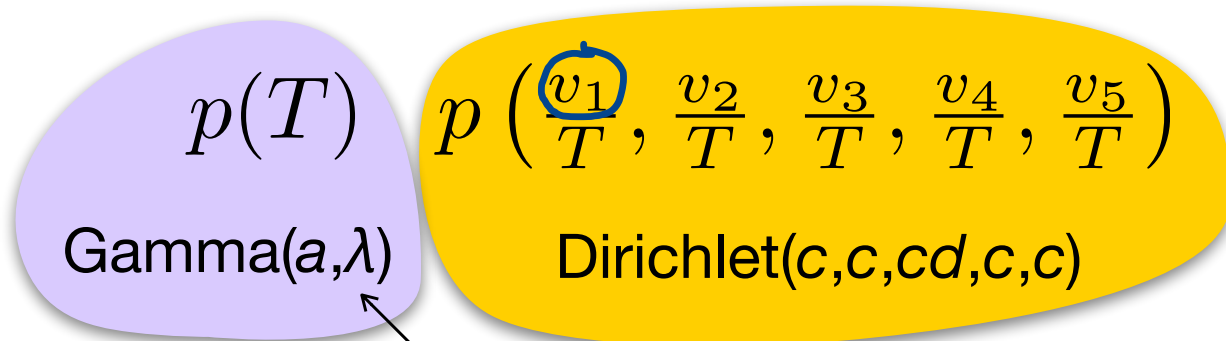


For tree with 100 edges, the induced tree length prior, Gamma(100, 0.1), has mean 10

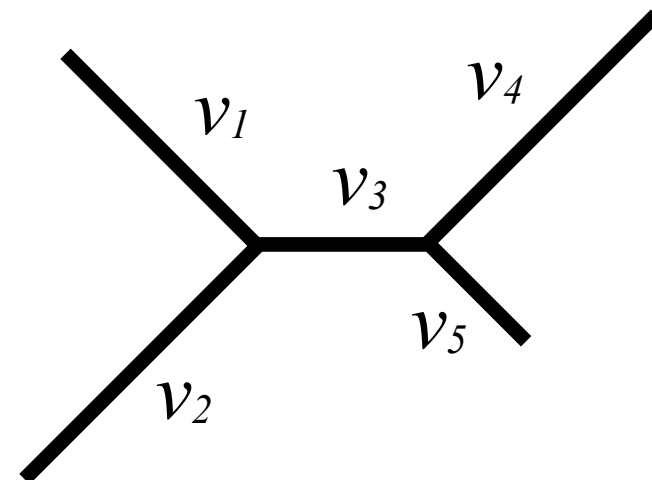


Gamma-Dirichlet prior

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

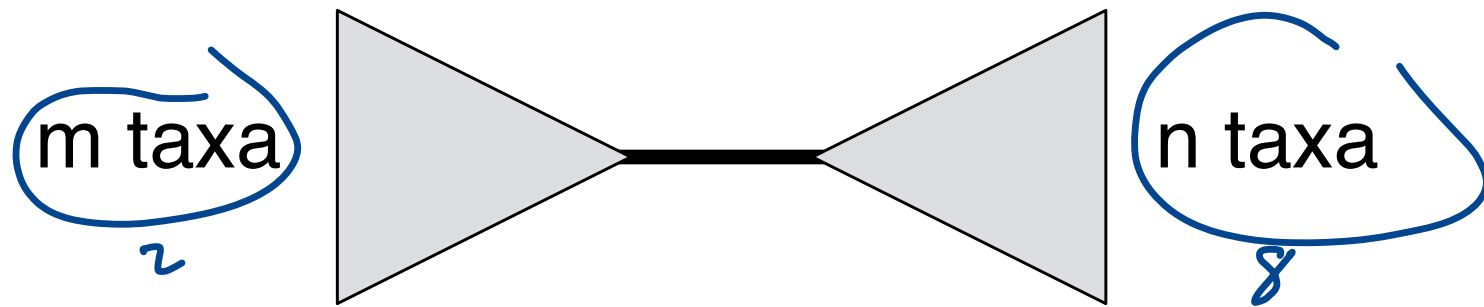


Note: *rate* parameterization: mean $T = a/\lambda$



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

Induced prior on splits



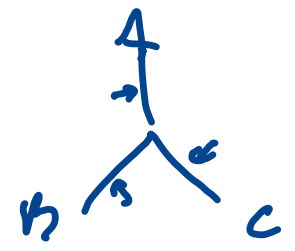
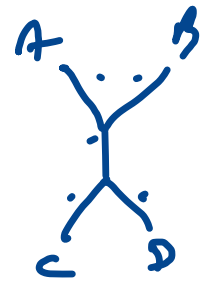
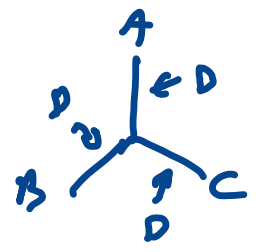
$$\text{Pr}(\text{split}) = \frac{\left[\begin{array}{c} \text{number of rooted} \\ \text{trees with } m \text{ taxa} \end{array} \right] \left[\begin{array}{c} \text{number of rooted} \\ \text{trees with } n \text{ taxa} \end{array} \right]}{\text{number of unrooted trees with } n+m \text{ taxa}}$$

202705

$m=2, n=8: \text{Pr}(\text{split}) = 0.0667$ ←

$m=5, n=5: \text{Pr}(\text{split}) = \underline{0.0001}$ ←

taxa	unrooted	rooted
2		1
3	1	3
4	$1 \cdot 3 = 3$	15
5	$3 \cdot 3 = 9$	105
6	$15 \cdot 7 = 105$	945
7	$105 \cdot 9 = 945$	10395
8	$945 \cdot 11 = 10395$	<u>135135</u>
9	$10395 \cdot 13 = 135135$	2027025
10	$135135 \cdot 15 = 2,027,025$	



Running on empty

What is the likelihood?

```
#NEXUS
```

```
begin data;
```

```
Dimensions ntax=4 nchar=1;
```

```
Format datatype=dna missing=?;
```

```
matrix
```

```
taxon1 ?
```

```
taxon2 ?
```

```
taxon3 ?
```

```
taxon4 ?
```

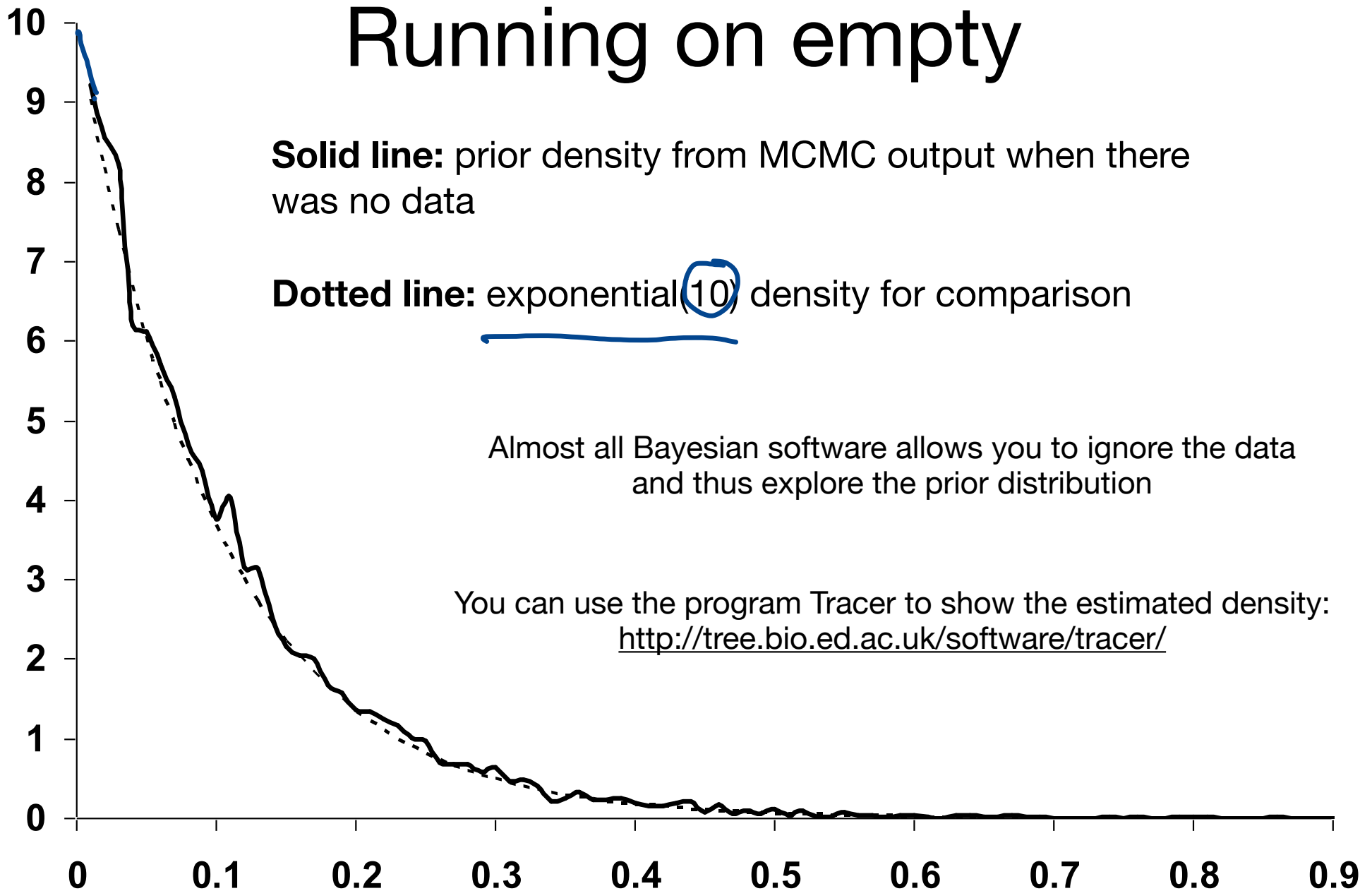
```
;
```

```
end;
```

Running on empty

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

Dotted line: exponential(10) density for comparison

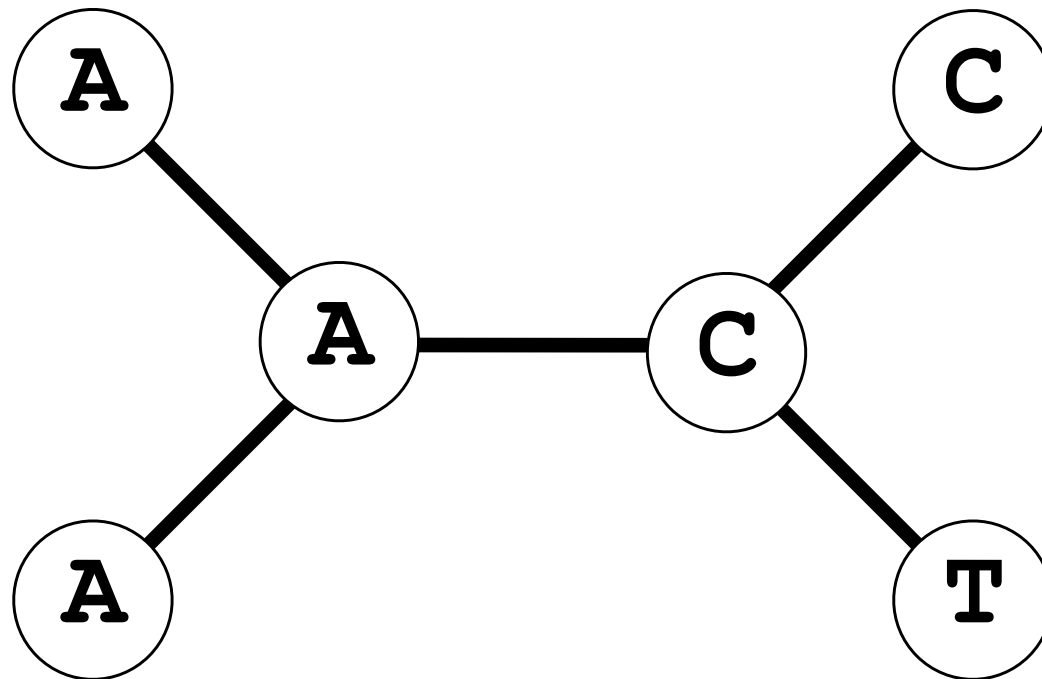


Hierarchical models vs. Empirical Bayes

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

Prior: Exponential, mean=0.1

$$L_k = \frac{1}{4} \left[\frac{1}{4} + \frac{3}{4} e^{-4v_1/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_2/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_3/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_4/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_5/3} \right]$$



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

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

hyperprior is the prior for the hyperparameter

Prior: Exponential, mean μ

$$L_k = \frac{1}{4} \left[\frac{1}{4} + \frac{3}{4} e^{-4v_1/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_2/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_3/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_4/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_5/3} \right]$$

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.

Empirical Bayes

uses data, hence empirical



Prior: Exponential, mean=MLE

$$L_k = \frac{1}{4} \left[\frac{1}{4} + \frac{3}{4} e^{-4v_1/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_2/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_3/3} \right] \left[\frac{1}{4} - \frac{1}{4} e^{-4v_4/3} \right] \left[\frac{1}{4} + \frac{3}{4} e^{-4v_5/3} \right]$$