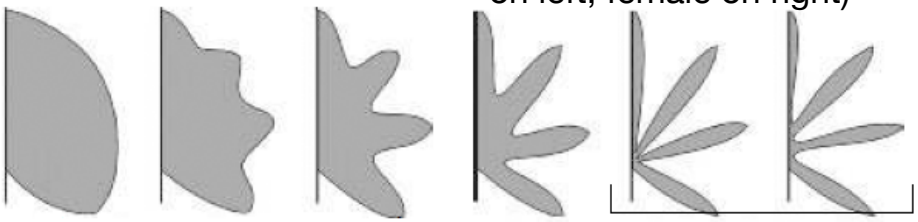
Discrete Morphology Models

Discrete morphological data

- Binary characters
  - e.g. wings: present (1) or absent (0)
  - e.g. venation: palmate (0) or pinnate (1)
- Multistate characters
  - e.g. Number of incisors: 0, 1, 2, 3, 4, 5
  - e.g. plants: hermaphroditic (0), gynomonoecious (1), or dioecious (2)
- Discretized continuous characters
  - e.g. snout-vent length: short (0) or long (1)
  - e.g. leaf shape: entire (0), lobed (1), deeply dissected (2)

Polygonella robusta is gynomonoecious (female flowers at tip, bisexual flowers at base)

Polygonella gracilis is dioecious (male individual on left, female on right)



Leaf lobing in *Pelargonium*. From online supplement to Jones et al. 2009. Evolution 63: 479-497

Symmetric vs. Asymmetric Models

	0	1
0	$-\beta$	β
1	β	$-\beta$

Symmetric
(forward rate equals reverse rate)

	0	1
0	$-\alpha$	α
1	β	$-\beta$

Asymmetric
(forward rate potentially differs from reverse rate)

Pagel 1994; Schluter et al. 1997

State Frequencies

	0	1
0	$-\alpha$	α
1	β	$-\beta$

← This instantaneous rate matrix...

...implies these equilibrium state frequencies:

$$\pi_0 = \frac{\beta}{\alpha + \beta} \qquad \pi_1 = \frac{\alpha}{\alpha + \beta}$$

In other words, if you know α and β, you also know π_0 and π_1, and if you know π_0 and π_1, you also know α and β

State Frequencies

Example: if rate of forward change (α, 0→1) is twice the rate of the reverse change (β, 1→0), then $\alpha = 2\beta$ and we have...

$$\pi_0 = \frac{\beta}{\alpha + \beta} = \frac{1}{3}$$

$$\pi_1 = \frac{\alpha}{\alpha + \beta} = \frac{2}{3}$$

In this case, fewer taxa are expected to have state 0 because of the greater tendency to change to 1

For the record...

Here are general formulas for the transition probabilities for the two-state model:

$$P_{00}(t) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t}$$

$$P_{01}(t) = \frac{\alpha}{\alpha+\beta}\left(1 - e^{-(\alpha+\beta)t}\right)$$

$$P_{10}(t) = \frac{\beta}{\alpha+\beta}\left(1 - e^{-(\alpha+\beta)t}\right)$$

$$P_{11}(t) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t}$$

Expected number of changes/site = $\frac{2\alpha\beta t}{\alpha+\beta}$

Multistate models

- Can extend the symmetric model to multiple states
  - 4-state version is identical to JC69
  - *k*-state version (where *k* is arbitrary) often called the Mk model (M=Markov)
- **Example of the model of evolution for a trait that adopts three states**

State	*0*	*1*	*2*
0	--	q_{01}	q_{02}
1	q_{01}	--	q_{12}
2	q_{20}	q_{21}	--

From the BayesTraits manual:

http://www.evolution.reading.ac.uk/Files/BayesTraits-V1.0-Manual.pdf

Lewis 2001

Molecules vs. Morphology

These two A nucleotides represent the **same state**

```
A ACCAT
B ACCAT
C ATCAT
D ATCAT
E ATCGT
```

These two 0 states have **nothing to do with each other**

```
A 00000
B 00000
C 10011
D 00011
E 10011
```

For morphology, it makes no sense to compute empirical frequencies

Lewis 2001

What to do about state frequencies?

1. Don't even try to estimate state frequencies

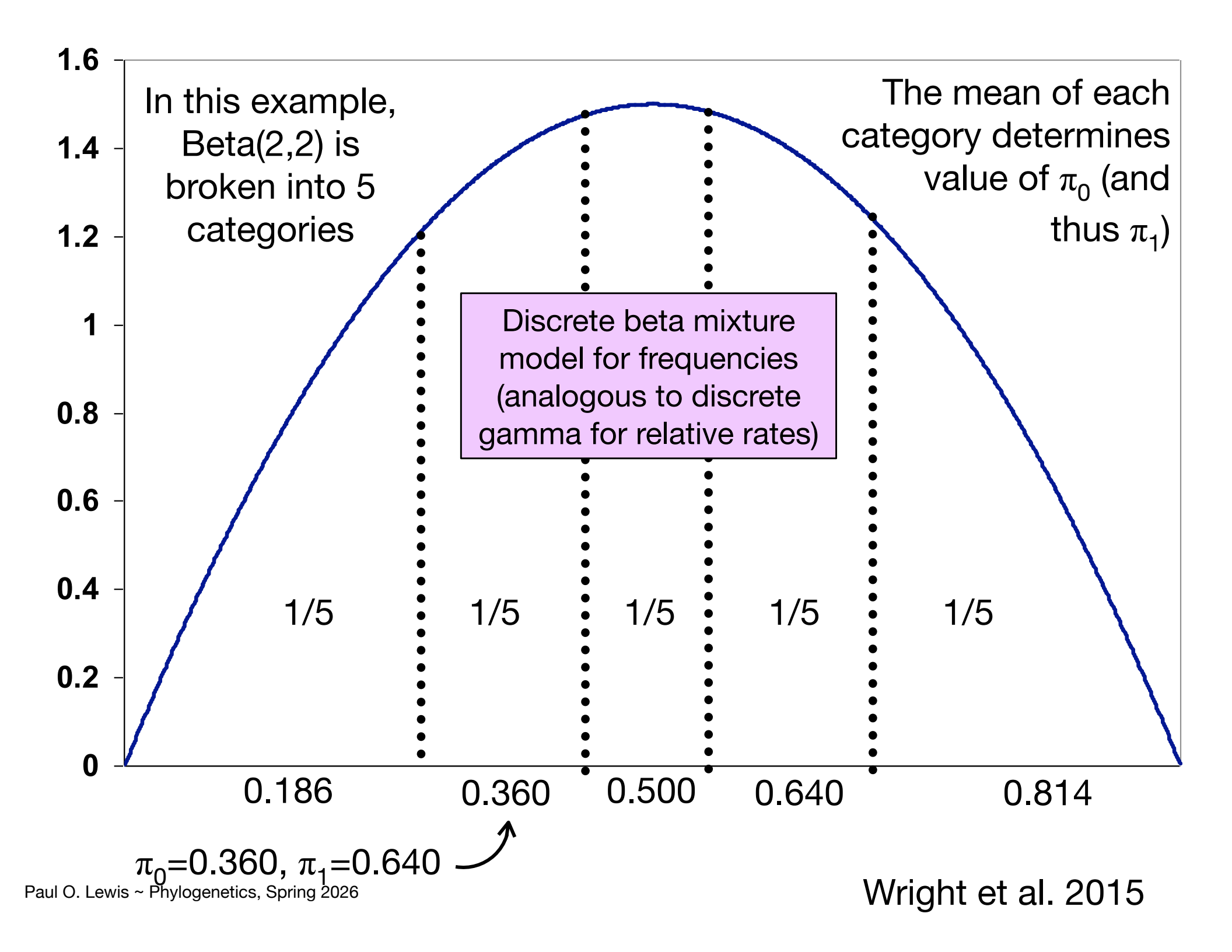
   Assume symmetric model; i.e. $\pi_0 = \pi_1 = 0.5$

2. Estimate state frequencies separately for each character

   Asymmetric model, but adds one parameter for every character

3. Use a mixture model

   Use **discrete beta distribution** for frequency heterogeneity in the same way that the **discrete gamma distribution** is used for rate heterogeneity


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
In this example, Beta(2,2) is broken into 5 categories
The mean of each category determines value of π_0 (and thus π_1)
Discrete beta mixture model for frequencies (analogous to discrete gamma for relative rates)
1/5
1/5
1/5
1/5
1/5
0.186
0.360
0.500
0.640
0.814
π_0=0.360, π_1=0.640
Wright et al. 2015

Ascertainment Bias in Morphology Datasets

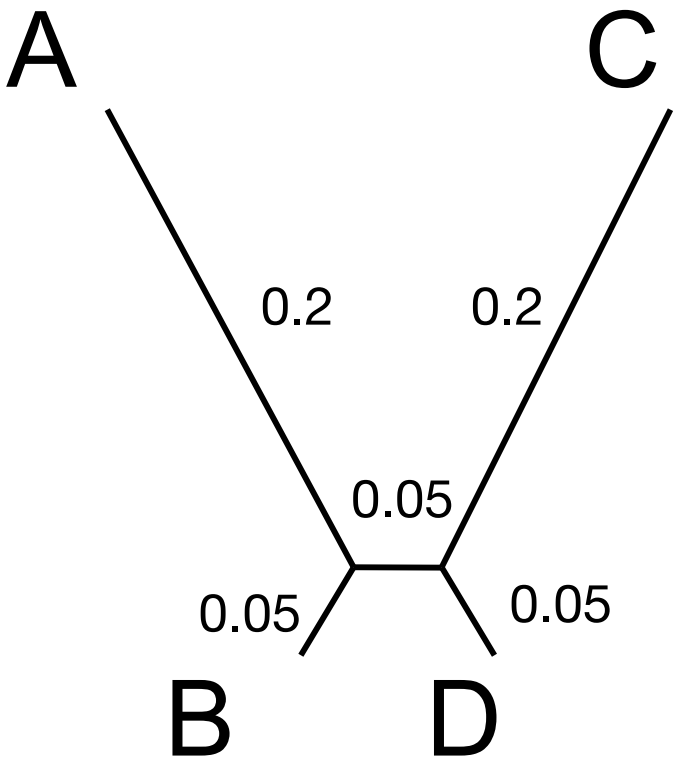
```
P._fimbriata      00000000110101010110?01
P._robusta        00000000110101010110001
P._articulata     10011101011000000l100
P._parksii        00011101211000000l000
P._americana      10011101011110000l000
P._myriophylla    10011101011100000l000
P._macrophylla    11011101210000000000
P._polygama       11011101310010000000
P._gracilis       10111101300000l000110
P._ciliata        00111011200000l000110
P._basiramia      00111011200000l000110
```

No constant characters: no characters have same state for all taxa

No autapomorphies (i.e. only one taxon different) either

This represents an **ascertainment bias**: characters included are biased towards those that are parsimony informative.

Estimating branch lengths from discrete morphological data



What if you simulated data using a JC69 model and this model tree, then *withheld all constant sites* from PAUP* and asked it to estimate branch lengths under the same JC69 model?

Lewis 2001

Estimating branch lengths from discrete morphological data

Doing that produces results like this:

Edge	True length	ML estimate
A	0.20	241750.00
B	0.05	0.43
C	0.20	54.65
D	0.05	143950.00
interior	0.05	0.02

Estimated edge lengths are crazy!

Conditioning on variability

$$\Pr(D, V) = \Pr(D|V)\Pr(V)$$

Probability of the data *and* character is variable

Probability of the data *given* character is variable

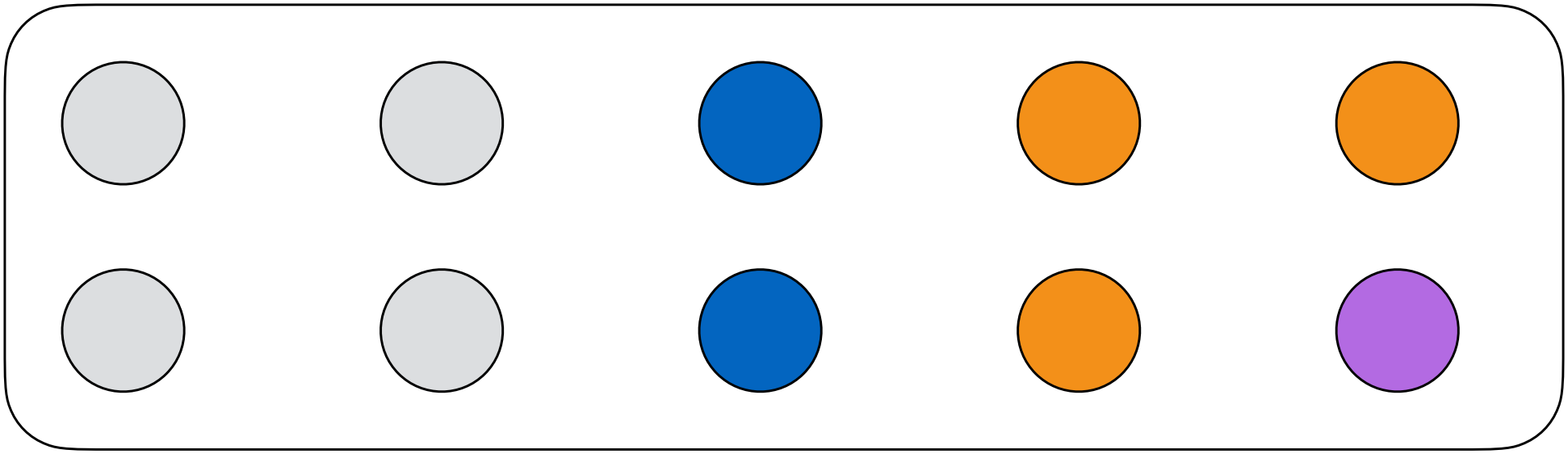
Probability that character is variable

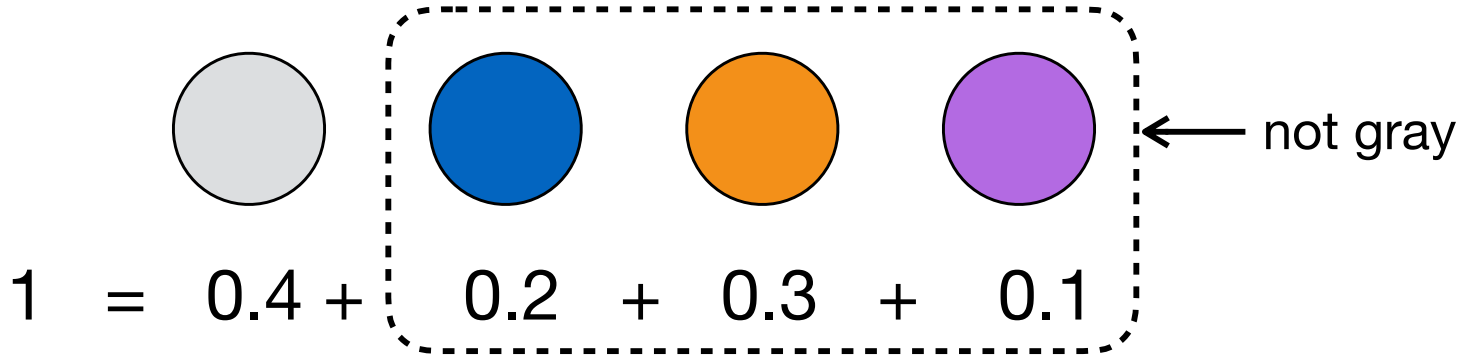
$$\Pr(D|V) = \frac{\Pr(D, V)}{\Pr(V)}$$

Likelihood computed the normal way

Likelihood conditional on character variability

Felsenstein 1992





Probability of choosing an **orange** circle = 0.3

Probability of choosing an **orange** circle **given** that the circle chosen is **not gray** = 0.3/0.6 = 0.5

Poisson Example

Let ***y*** be the **number of accidents** at an intersection/week.
Lambda (λ) is the **mean** number of accidents/week.

$$1 = \overset{0}{\frac{\lambda^0 e^{-\lambda}}{0!}} + \overset{1}{\frac{\lambda^1 e^{-\lambda}}{1!}} + \overset{2}{\frac{\lambda^2 e^{-\lambda}}{2!}} + \overset{3}{\frac{\lambda^3 e^{-\lambda}}{3!}} + \cdots$$

These are the probabilities of 0, 1, 2, 3, ..., accidents/week given lambda (the infinite sum equals 1, as it should).

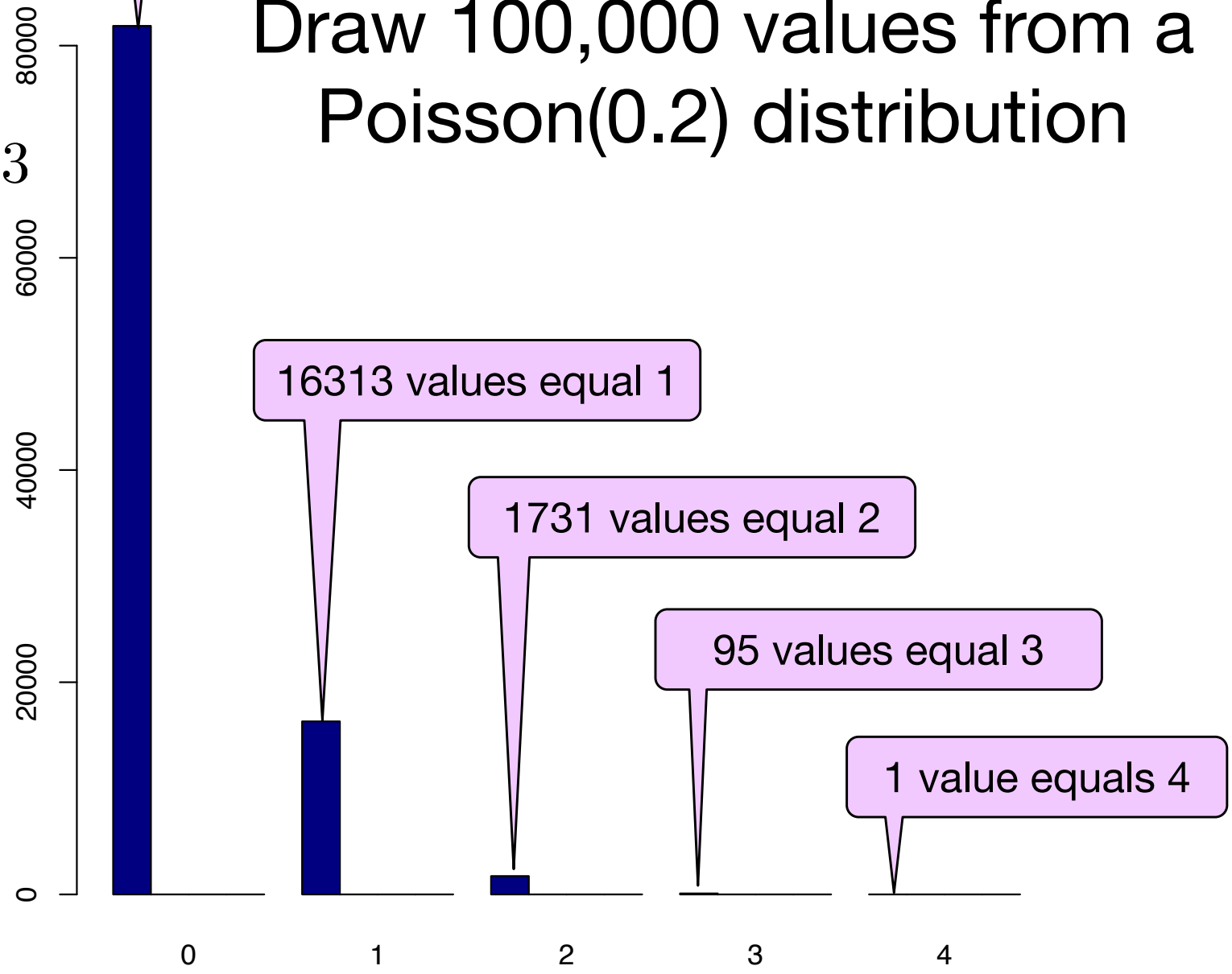
Draw 100,000 values from a Poisson(0.2) distribution

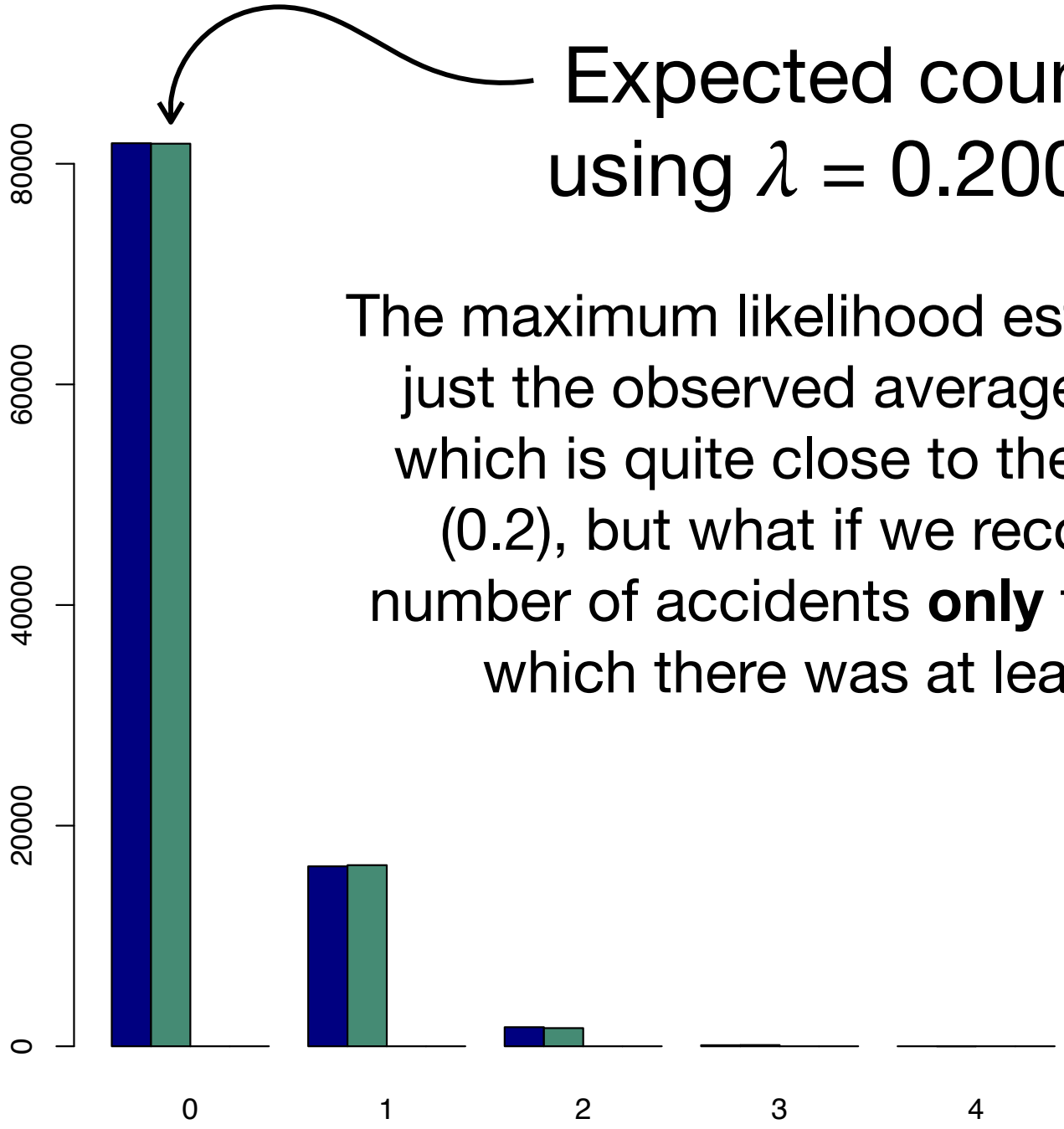
81860 values equal 0

good match

$e^{-\lambda} = 0.81873$

There is an 82% chance of having zero accidents/ week

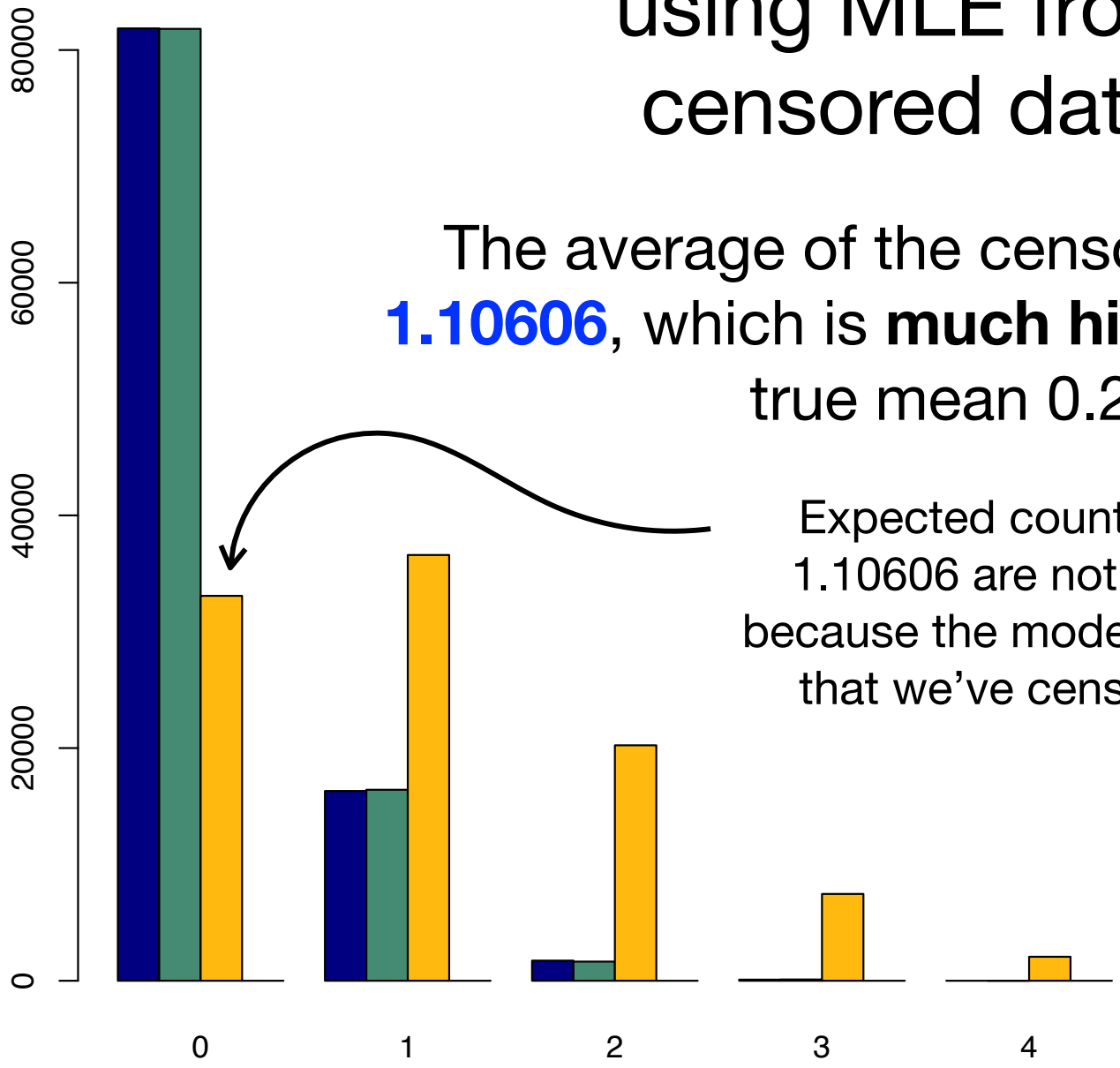




The maximum likelihood estimate of λ is just the observed average, 0.20064, which is quite close to the true value (0.2), but what if we recorded the number of accidents **only** for weeks in which there was at least one?

Expected counts using MLE from censored data

The average of the censored data is **1.10606**, which is **much higher** than the true mean 0.2



Expected counts using mean 1.10606 are not very accurate because the model does not know that we've censored the data

Poisson Example

$$1 = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \cdots$$

$$1 - e^{-\lambda} = \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \cdots$$

We should be using this as our total probability

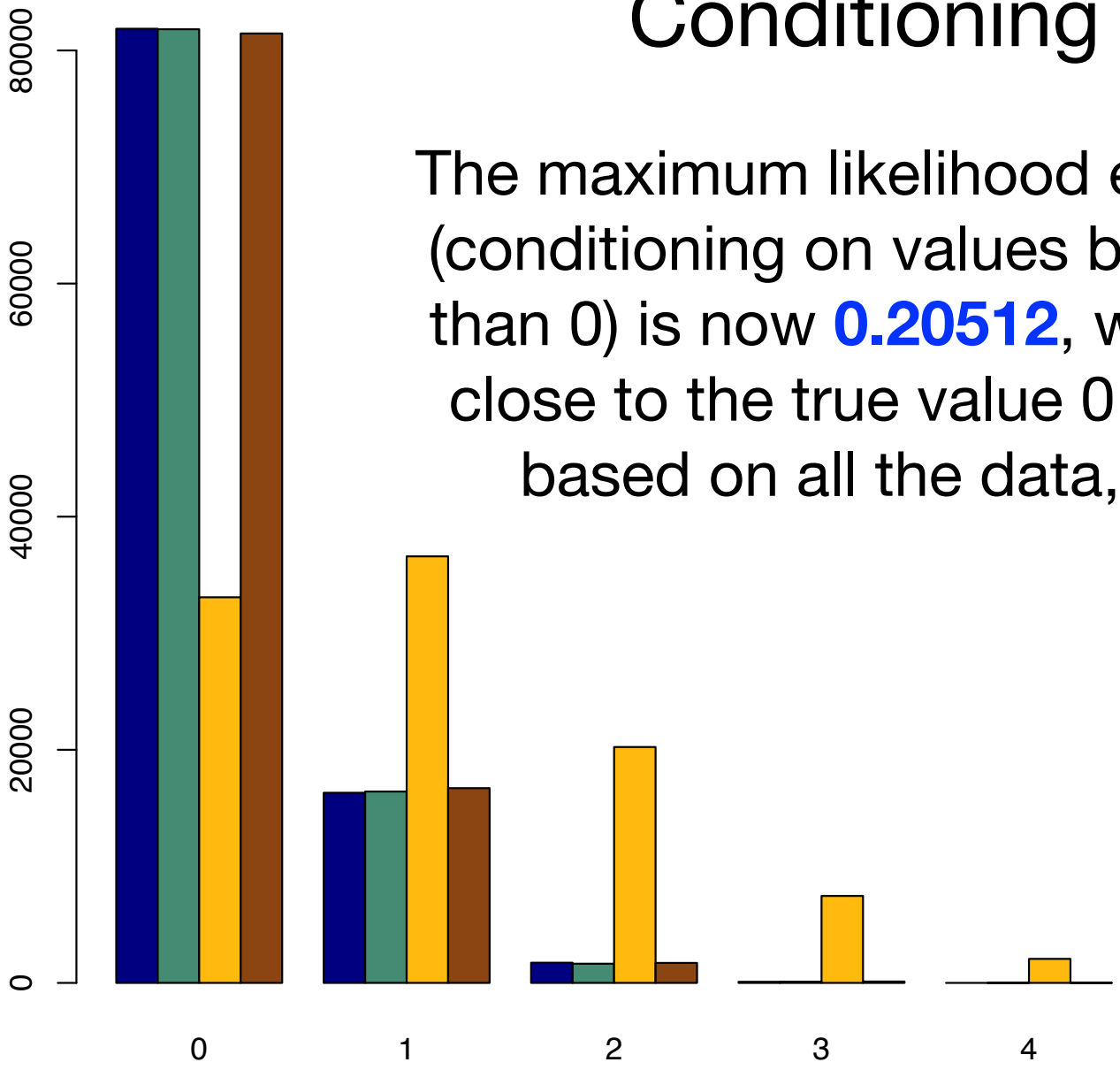
Poisson Example

1 2 3

$$1 = \frac{\lambda^1 e^{-\lambda}}{1!(1-e^{-\lambda})} + \frac{\lambda^2 e^{-\lambda}}{2!(1-e^{-\lambda})} + \frac{\lambda^3 e^{-\lambda}}{3!(1-e^{-\lambda})} +$$

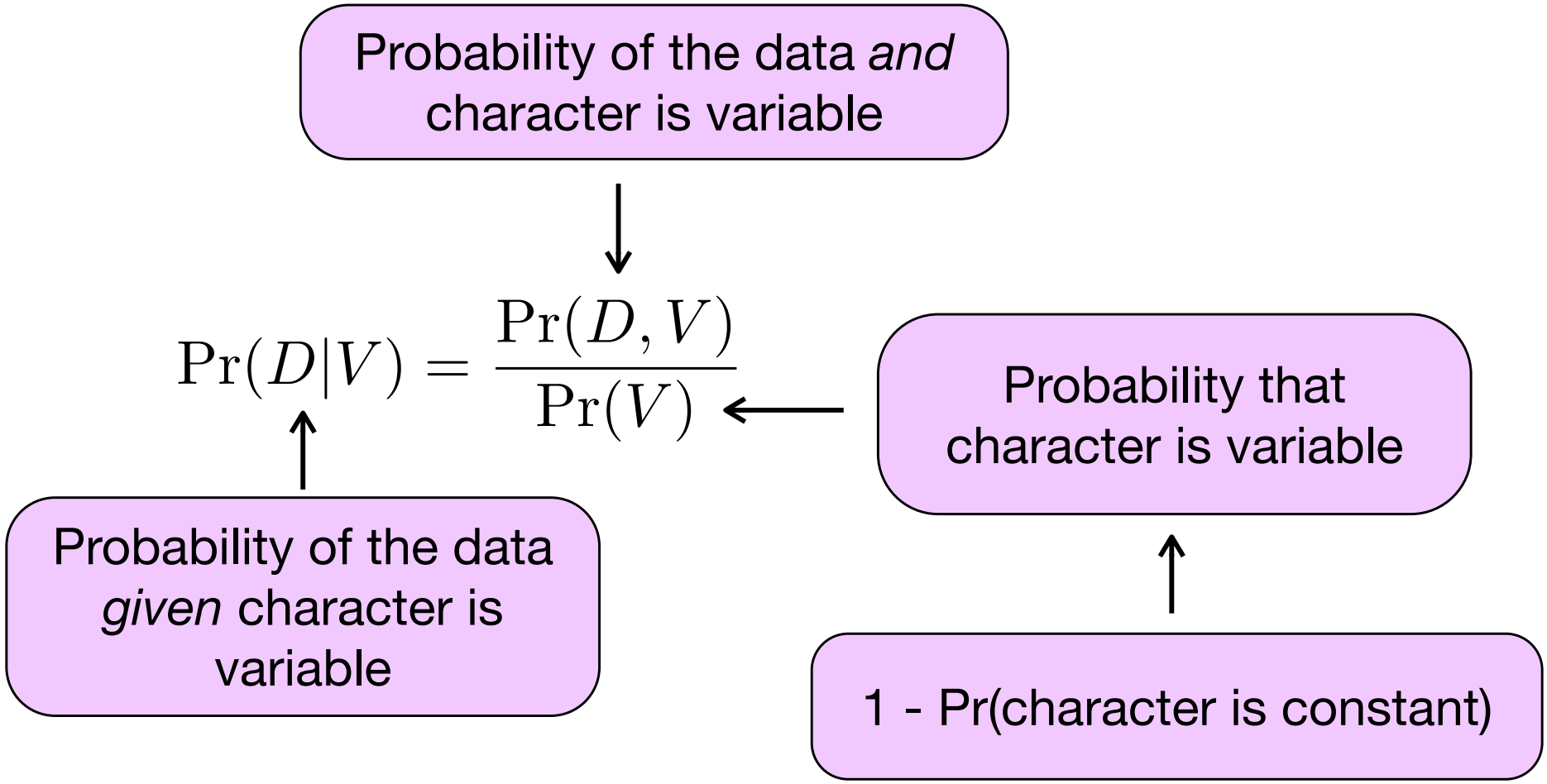
Dividing each by the probability of at least one accident serves to inform the model that we've omitted the zeros

Conditioning helps!



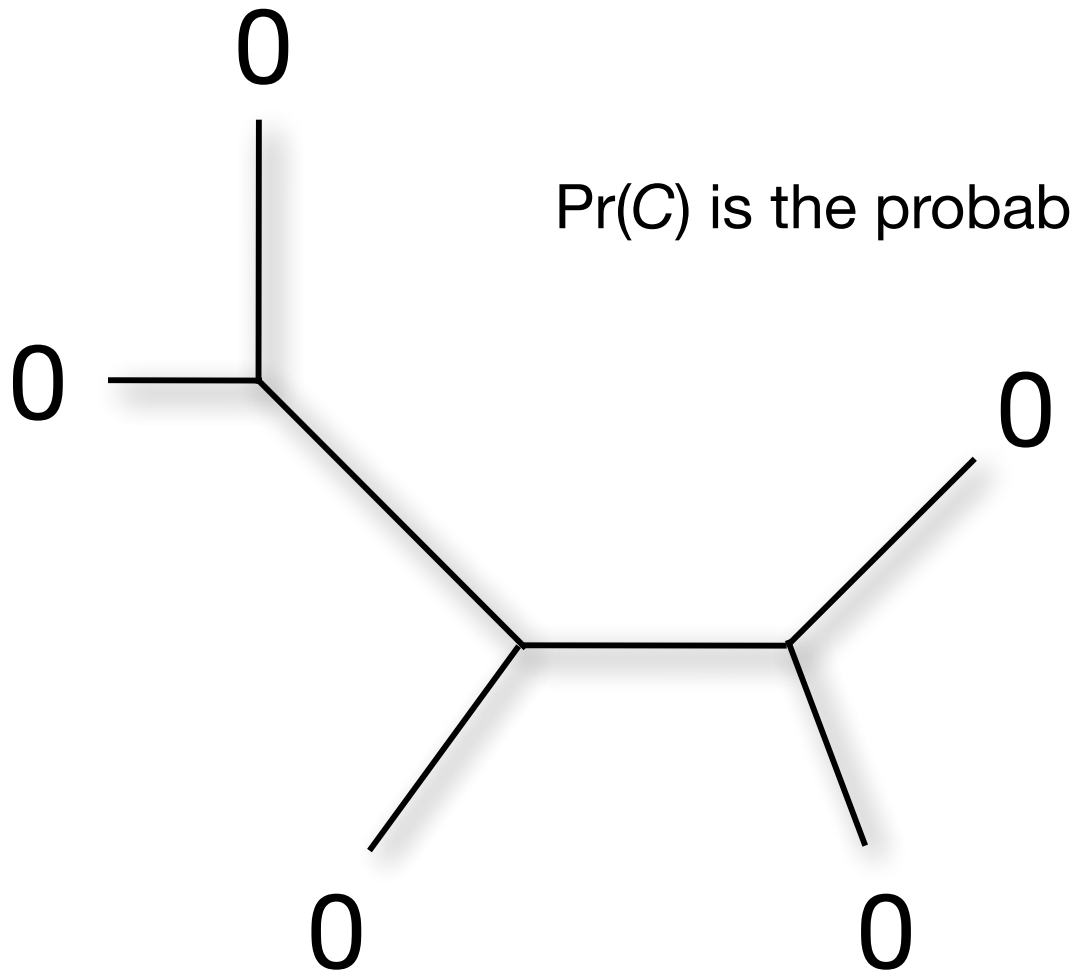
The maximum likelihood estimate of λ (conditioning on values being greater than 0) is now **0.20512**, which is very close to the true value 0.2 and MLE based on all the data, 0.20064

Conditioning on variability



How do we calculate Pr(C), the probability that a character is constant?

Calculating Pr(*C*)



Pr(*C*) is the probability that a constant character like this one would arise based on the current tree topology and edge lengths.

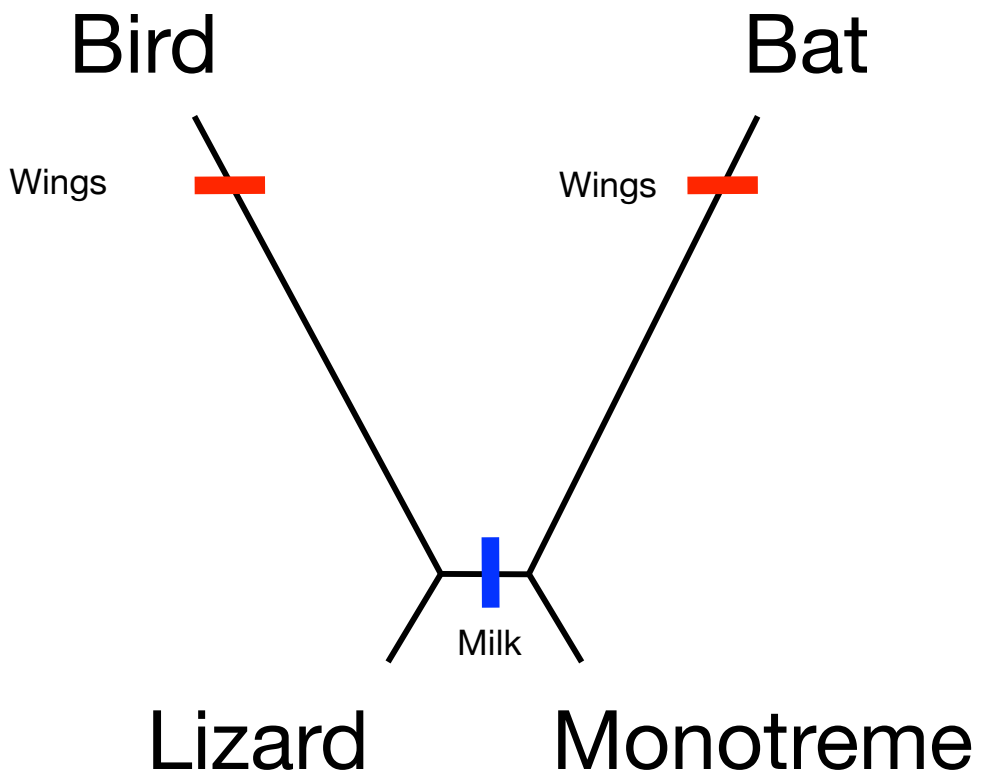
Estimating branch lengths from discrete morphological data

Here is the result of conditioning on variability:

Edge	True length	Naïve model	Corrected model
A	0.20	241750.00	0.21
B	0.05	0.43	0.05
C	0.20	54.65	0.21
D	0.05	143950.00	0.05
interior	0.05	0.02	0.05

Much better!

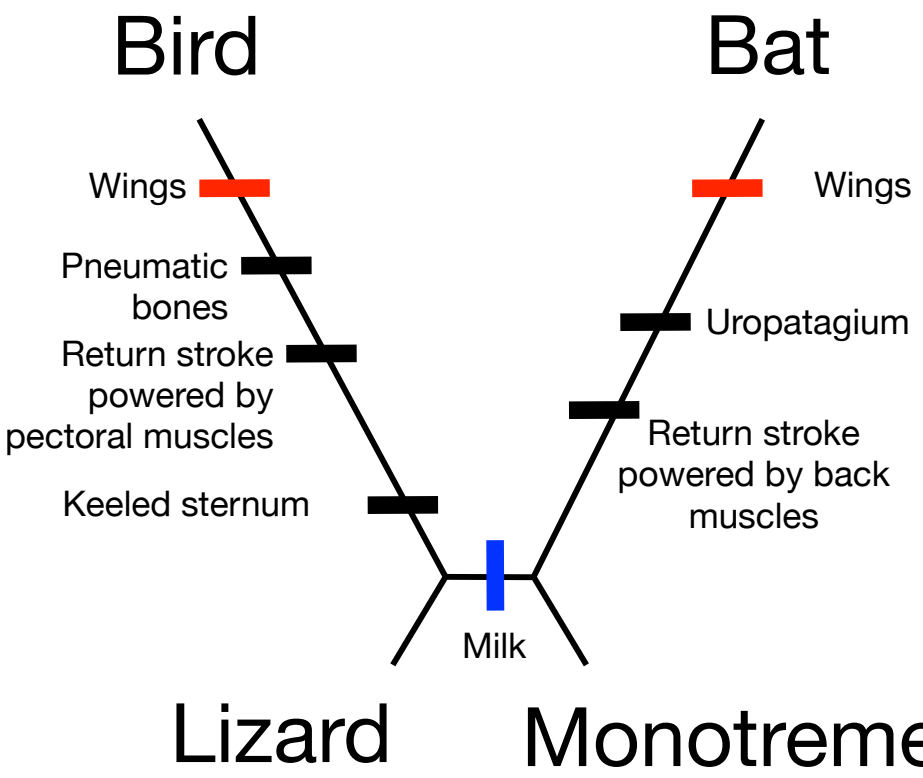
The Autapomorphy Trail



Wings and **milk** are both parsimony informative, but conflict - one must be homoplasious. Parsimony would not be able to decide between the tree supported by wings and the alternative tree (shown here) supported by milk.

Lewis 2001

The Autapomorphy Trail



We know that wings is the homoplasy here, and both birds and bats evolved wings independently to allow them to fly.

Evidence for this independent adaptation lies in the **trail of autapomorphies** related to flight.

If convergence events are often associated with such a trail of autapomorphies, then using branch length information is helpful.

The no-common-mechanism model

Separate edge length parameter for every edge **and** every site.

$$\max\{L\} = \left(\frac{1}{r}\right)^{s+1}$$

s = **parsimony tree length**
r = **number of states**

Minimizing the parsimony score (s) maximizes the likelihood.

It has been called the **no common mechanism** model because it decouples the evolution of separate characters: a character can do whatever it wants on any branch under this model.

This differs from the Mk model, for which every branch is associated with one length parameter that applies to all characters.

Tuffley and Steel (1997)

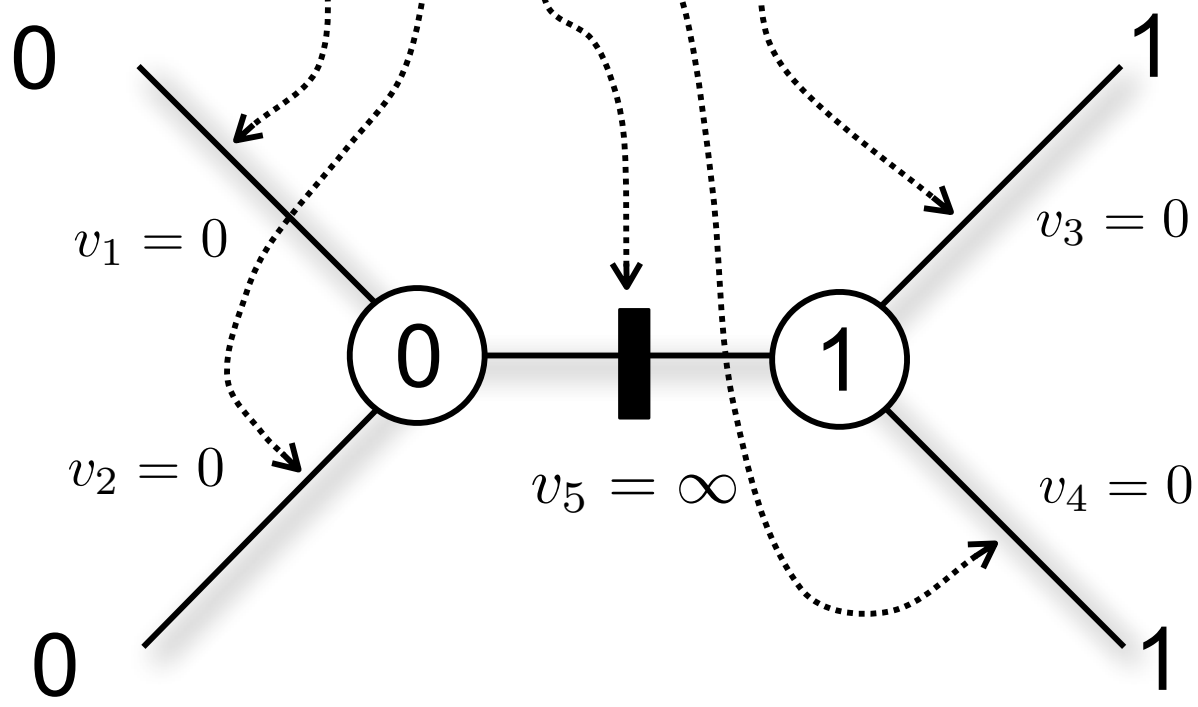
Which model fits better?
Which model makes better predictions

	T	H	H	T	H	T	T
M_1	$1-p$	p	p	$1-p$	p	$1-p$	$1-p$
M_2	$1-p_1$	p_2	p_3	$1-p_4$	p_5	$1-p_6$	$1-p_7$

Too good to be true?

Initial 1/2 is the probability of the starting state
Transition probabilities either 1 or 1/2 depending on whether the branch length is 0 or infinity, respectively

$$\max\{L\} = \left(\tfrac{1}{2}\right)(1)(1)\left(\tfrac{1}{2}\right)(1)(1) = \left(\frac{1}{2}\right)^2$$



Holder et al. 2010