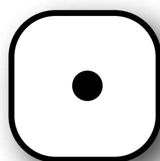
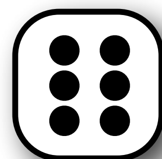
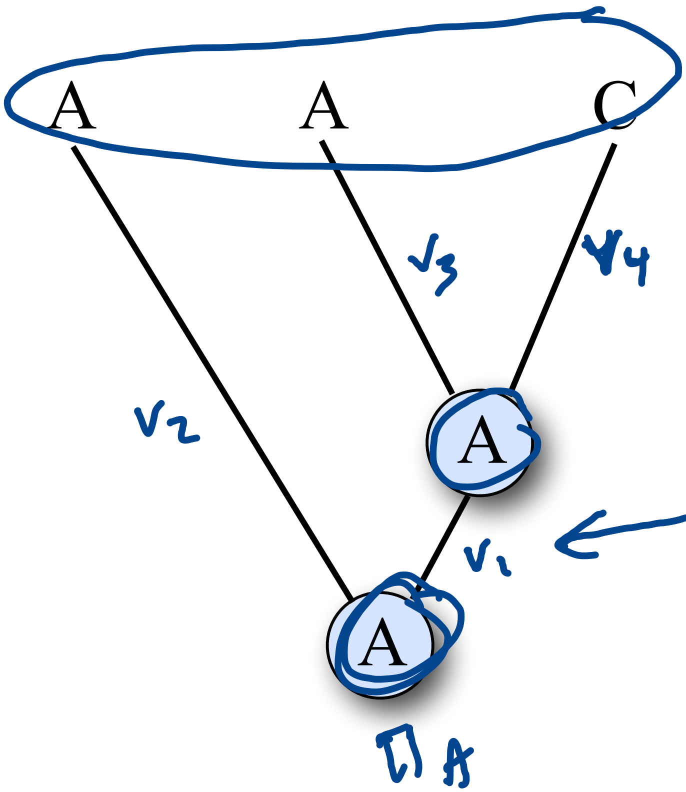
AND means MULTIPLY

Using 2 dice, what is the probability of

 AND  ?

$\left(\frac{1}{6}\right) \qquad \left(\frac{1}{6}\right) \quad = \frac{1}{36}$

AND rule in phylogenetics

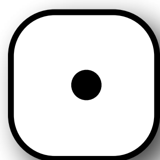
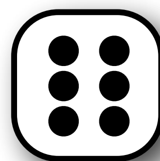


$$P(AACAA \mid \text{tree})$$

$$= \pi_A \, \underline{P_{AA}(v_1)} \, \underline{P_{AA}(v_2)} \, \underline{P_{AA}(v_3)} \, \underline{P_{AC}(v_4)}$$

OR means ADD

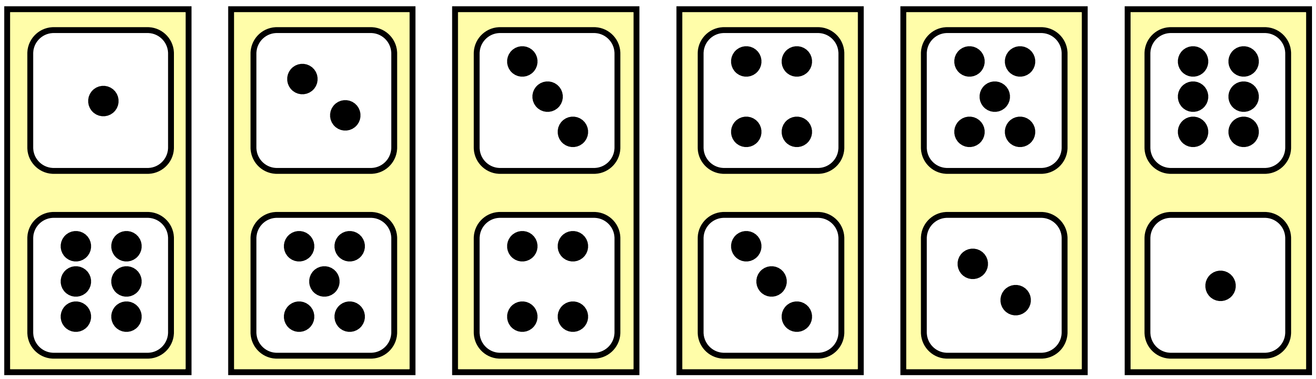
Using one die, what is the probability of

 OR  ?

$\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

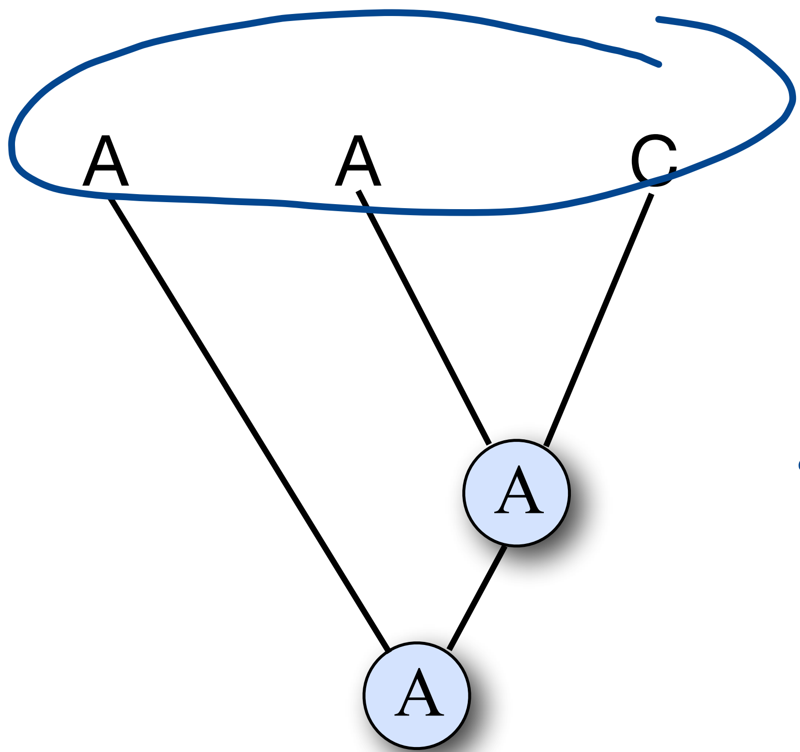
Combining AND and OR

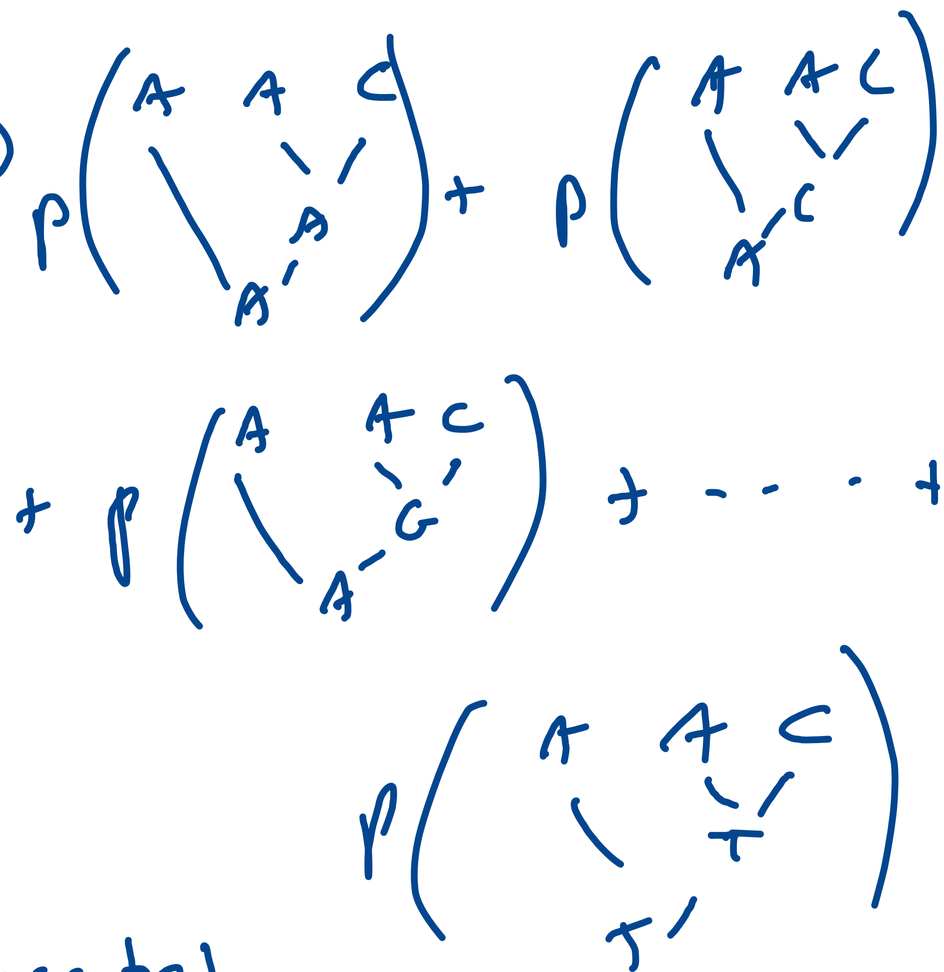
What is the probability that the sum of two dice is 7?



$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

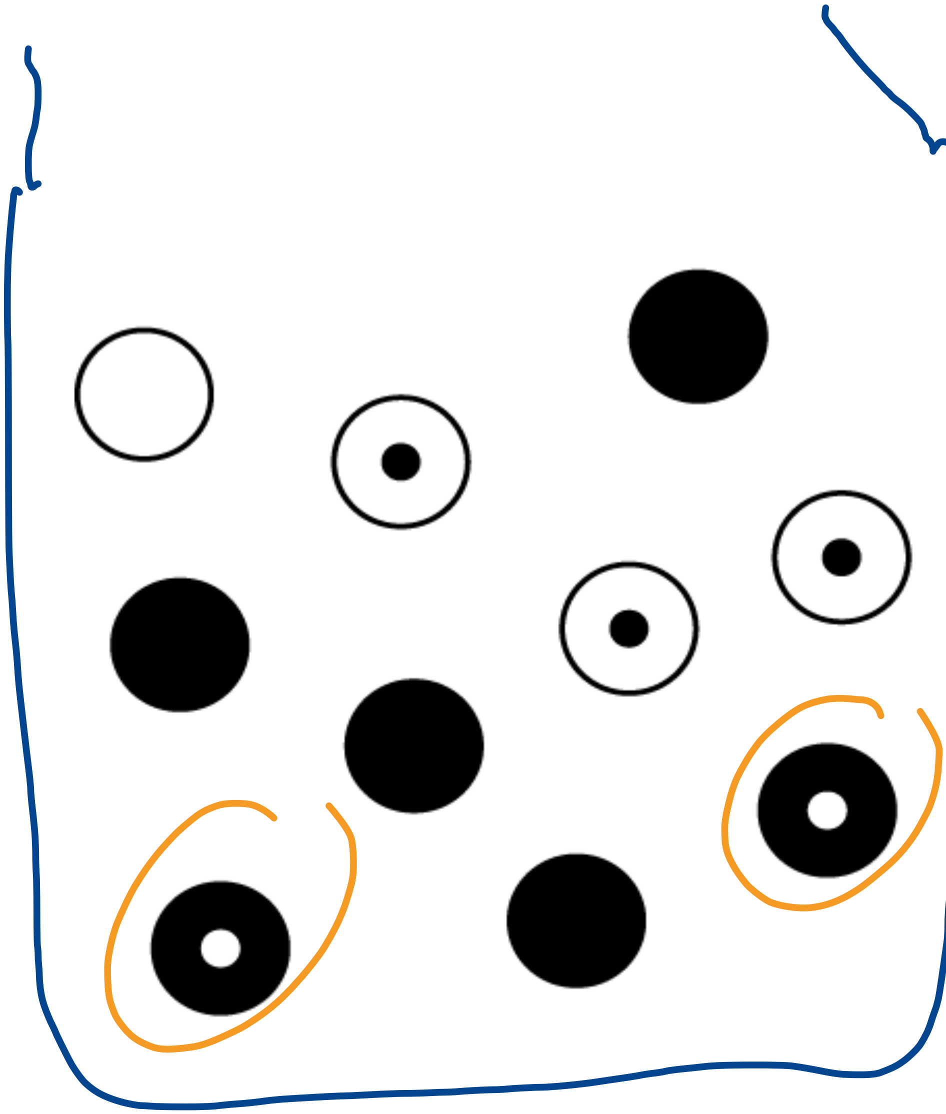
Using both AND and OR in phylogenetics





16 ancestral state combinations to sum across

Joint probabilities



B	=	**Black**	**S**	=	**Solid**
W	=	**White**	**D**	=	**Dotted**

Pr(B) = 0.6 Pr(S) = 0.5 solid

Pr(W) = 0.4 Pr(D) = 0.5 dotted

joint

Pr(◉) = Pr(B, D) = 0.2

Pr(●) = Pr(B, S) = 0.4

Pr(⊙) = Pr(W, D) = 0.3

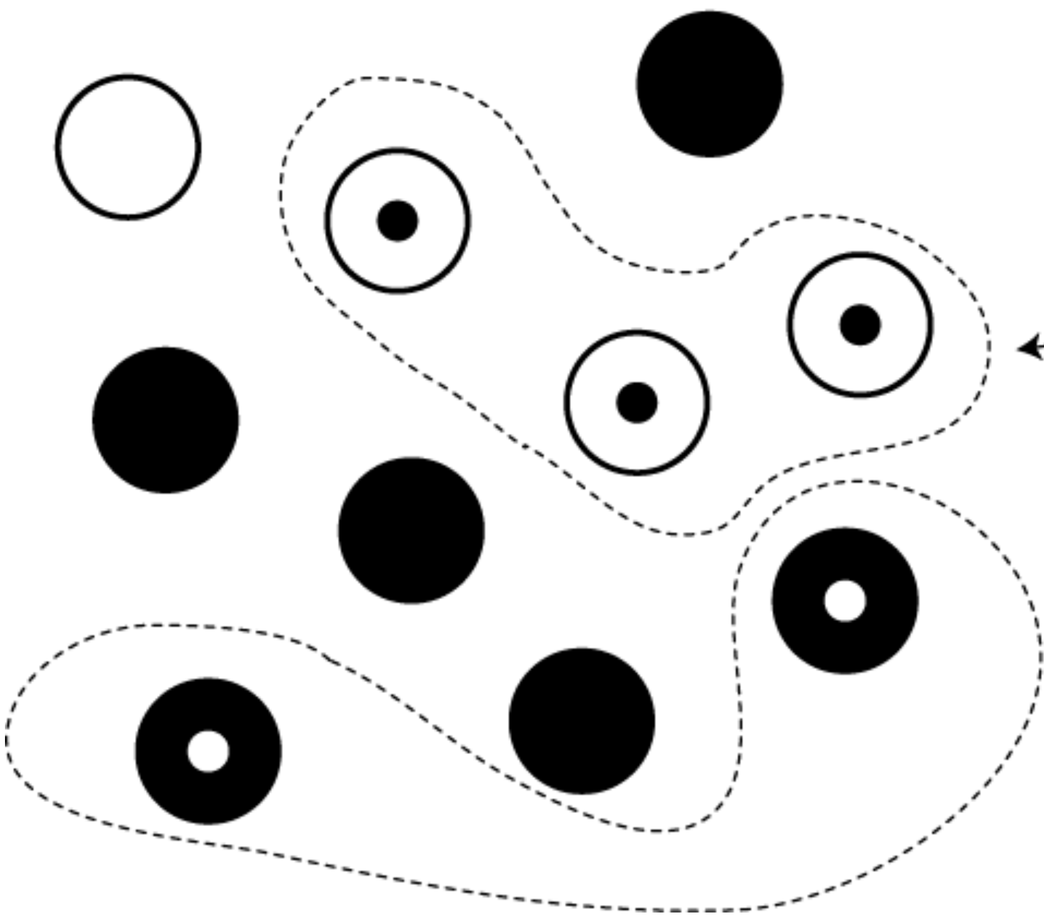
Pr(○) = Pr(W, S) = 0.1

Total probability of "Dotted"

$$\Pr(D) = \Pr(D,W) + \Pr(D,B)$$

.3 + .2

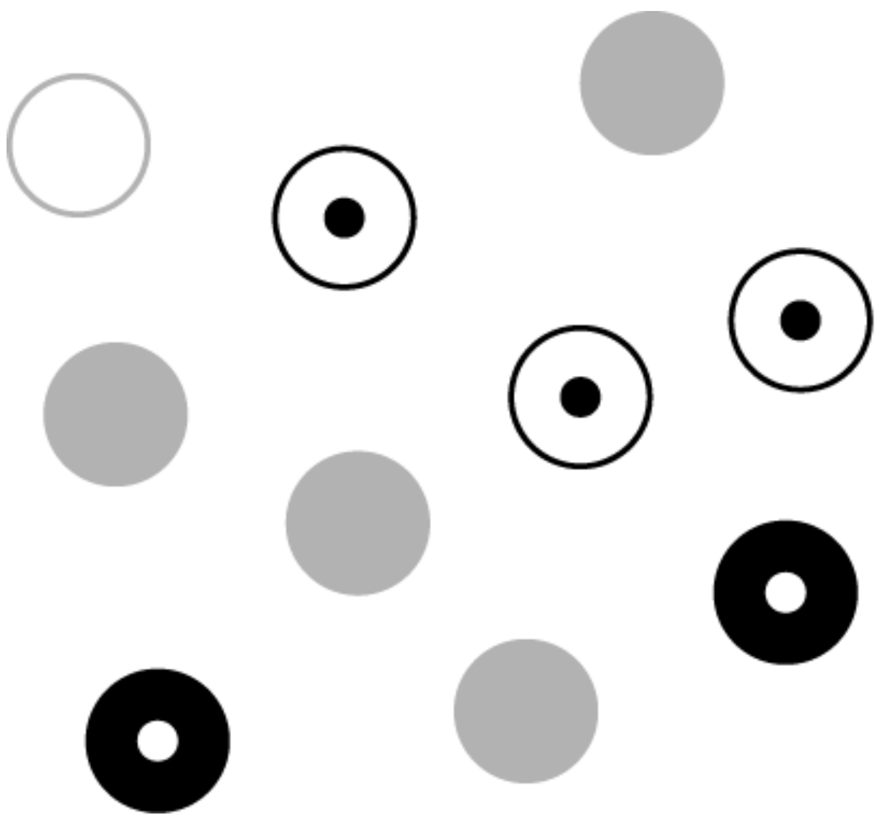
= .5



marginal probability of "dotted"

marginalizing over color

Conditional probabilities



condition is "dotted"

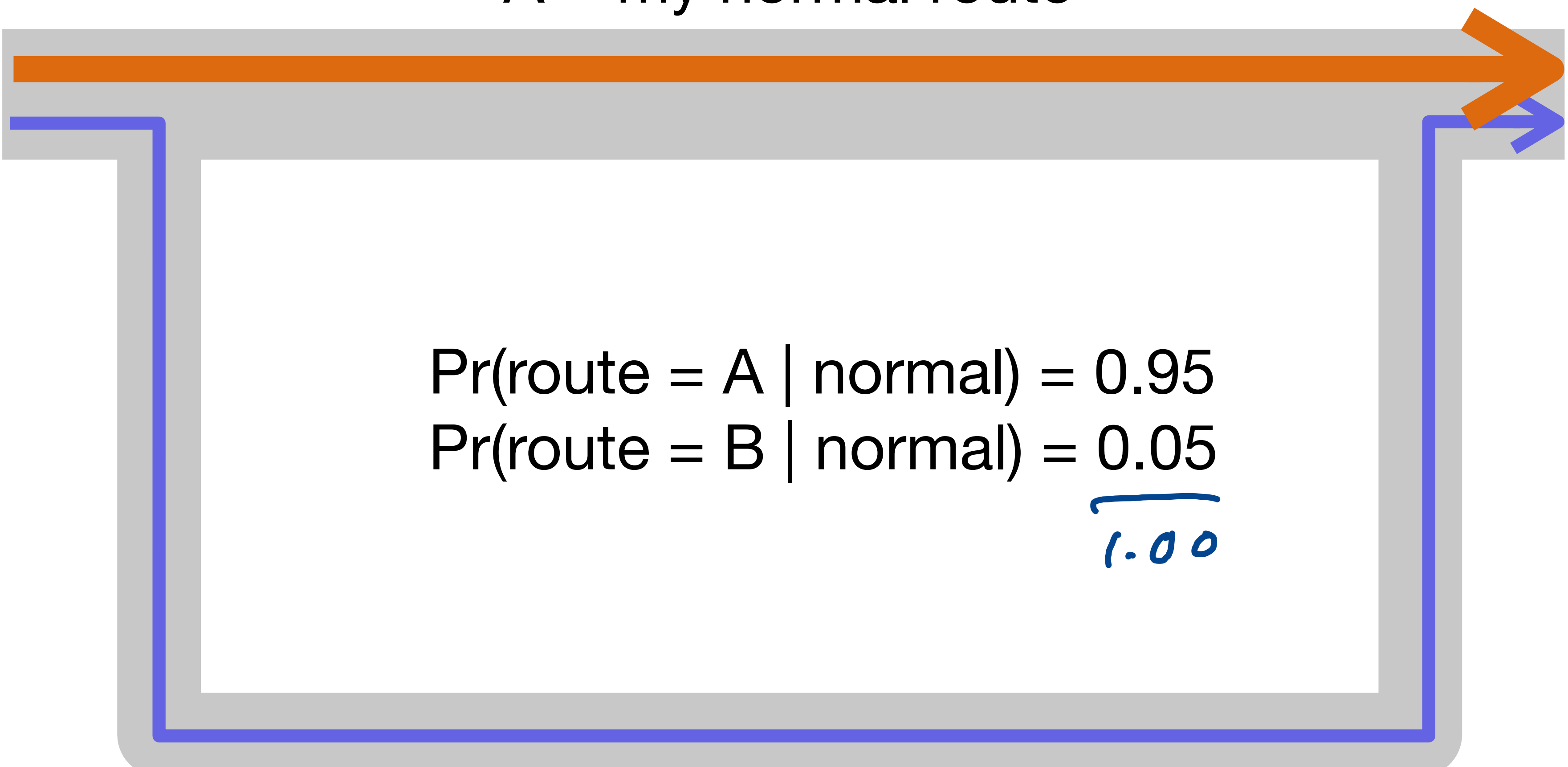
$$\Pr(B|D) = \frac{2}{5} = 0.4$$

Hide all solid marbles
(leaving 5 with dot)

Of those left, 2 are black

Dependence Example

A = my normal route



Pr(route = A | normal) = 0.95

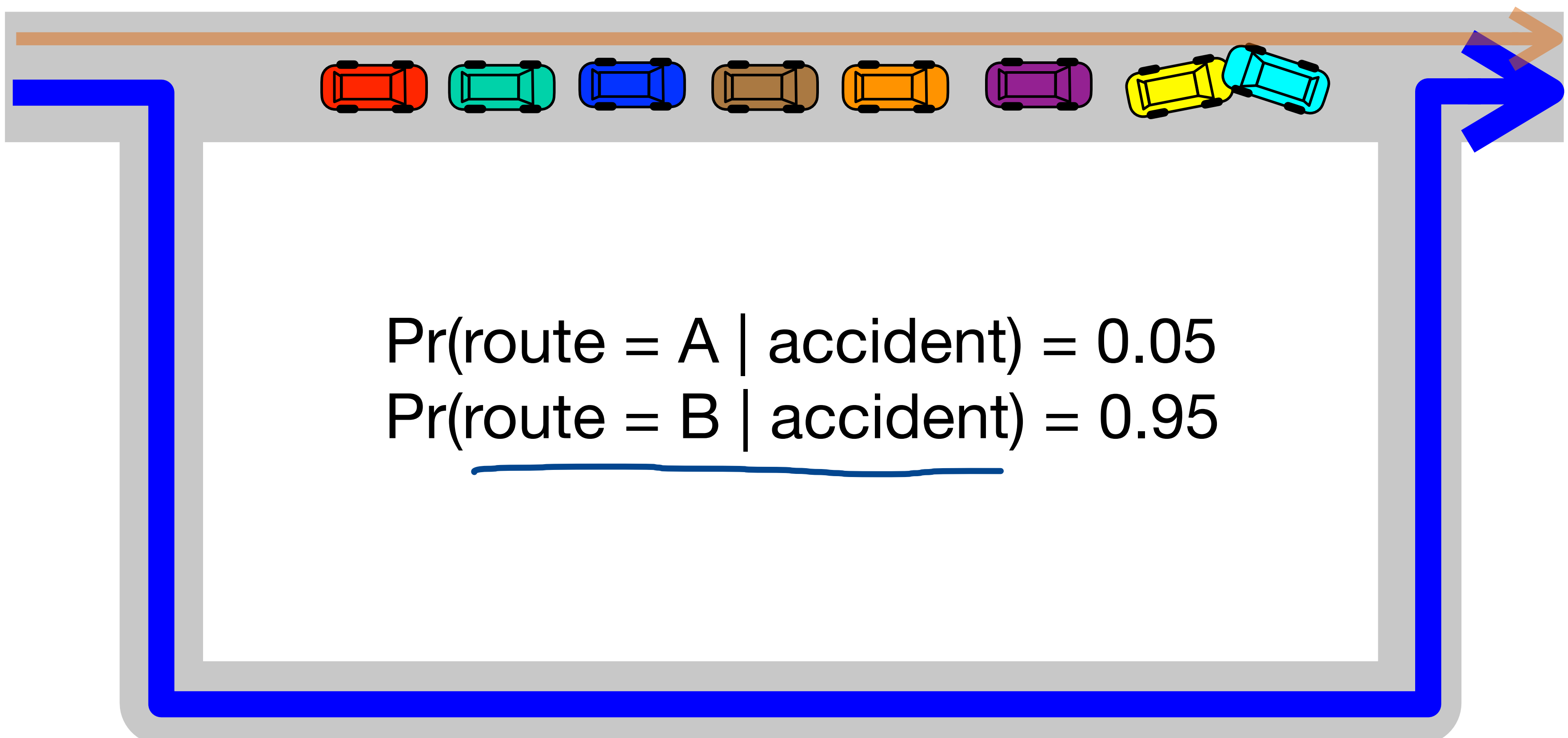
Pr(route = B | normal) = 0.05

1.00

B = I go out of my way

Dependence Example

A = my normal route



$\Pr(\text{route} = A \mid \text{accident}) = 0.05$

$\Pr(\text{route} = B \mid \text{accident}) = 0.95$

B = I go out of my way

How frequently do I take route B?

(assume Pr(accident) = 0.1)

$p(\text{normal}) = .9 \qquad p(\text{accident}) = .1$

$p(B|\text{normal}) = .05 \qquad p(B|\text{accident}) = .95$

$$p(B) = \underline{p(B, \text{normal})} + \underline{p(B, \text{accident})}$$

$$p(B|\text{normal})\,p(\text{normal}) + p(B|\text{accident})\,p(\text{accid.})$$

$$(.05)(.9) + (.95)(.1)$$

$$= .045 + .095$$

$$= .14 \qquad p(A) = .86$$

GAAGT ... GG

time

Likelihood of a single sequence

First 32 nucleotides of the ψη-globin gene of gorilla:

GAAGTCCTTGAGAAATAAACTGCACACACTGG

$$L_{JC} = \frac{1}{4}\frac{1}{4}\frac{1}{4} \cdots \frac{1}{4} = \left(\frac{1}{4}\right)^{32}$$ ← JC69 model

← F81 model

$$L_{F81} = \pi_G \pi_A \pi_A \pi_G \cdots \pi_G = \underline{\pi_A^{12} \pi_C^{7} \pi_G^{7} \pi_T^{6}}$$

$$\log L_{F81} = \underline{12 \log(\pi_A) + 7 \log(\pi_C) + 7 \log(\pi_G) + 6 \log(\pi_T)}$$

$$\log(ab) = \log(a) + \log(b)$$

$$\hat{\pi}_A = \frac{12}{32}$$

Likelihood ratio test

Find *maximum* logL under F81 (unconstrained) model:

-43.1 $\pi_A = .375$ $\pi_C = .219$ $\pi_G = .219$ $\pi_T = .187$

3 parameters estimated

Find *maximum* logL under JC69 (constrained) model:

-44.4 $\pi_A = \pi_C = \pi_G = \pi_T = 1/4$

1.3

0 parameters estimated

-43.1: .375, .219, .219, .187

-44.4: .25, .25, .25, .25

Likelihood ratio test

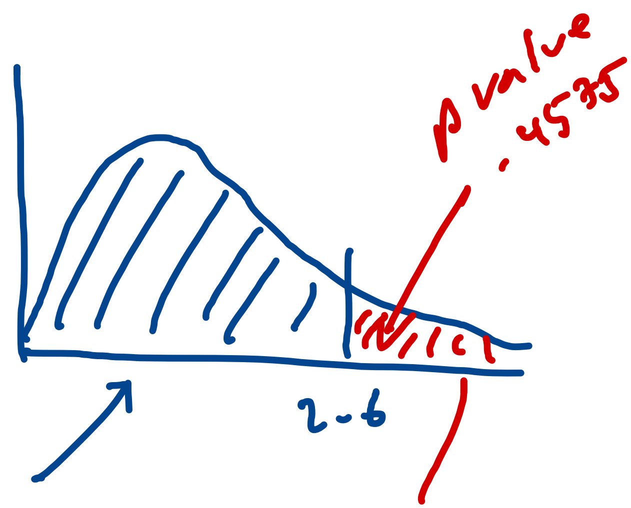
Calculate the likelihood ratio test statistic:

$$LRT = -2\left(\log L_{JC} - \log L_{F81}\right)$$

$$= 2.6$$

Calculate the degrees of freedom:

$$d.f. = 3 - 0 = 3$$



$$\text{pchisq}(\underset{\text{statistic}}{2.6}, \underset{\text{d.f.}}{3}) = \underline{.5425}$$

$$1 - .5425$$

$$= .4575$$

$$\text{qchisq}(.95, 3) = 7.814$$

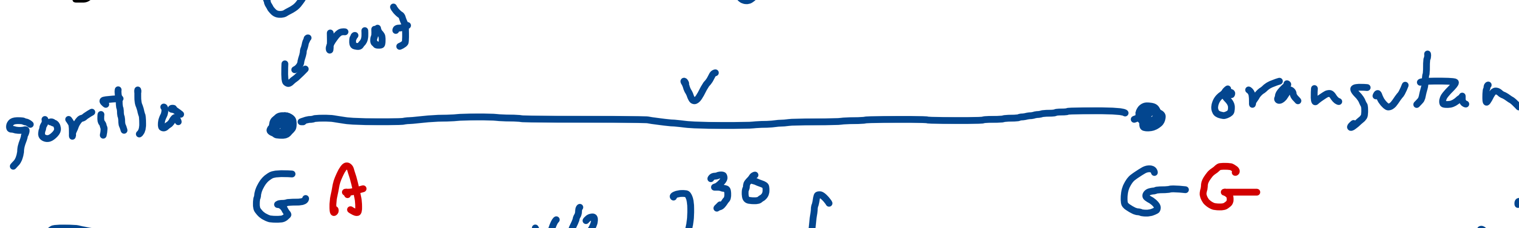
Likelihood of the simplest tree

First 32 nucleotides of the ψη-globin gene of gorilla and orangutan:

```
gorilla    GAAGTCCTTGAGAAATAAACTGCACACACTGG
orangutan  GGACTCCTTGAGAAATAAACTGCACACACTGG
```

$v = 3\beta t$

$\beta t = \frac{v}{3}$



$$L = \left[\left(\frac{1}{4}\right)\left(\frac{1}{4} + \frac{3}{4} e^{-4\cancel{\beta t}\, v/3}\right)\right]^{30} \left[\left(\frac{1}{4}\right)\left(\frac{1}{4} - \frac{1}{4} e^{-4v/3}\right)\right]^{2}$$

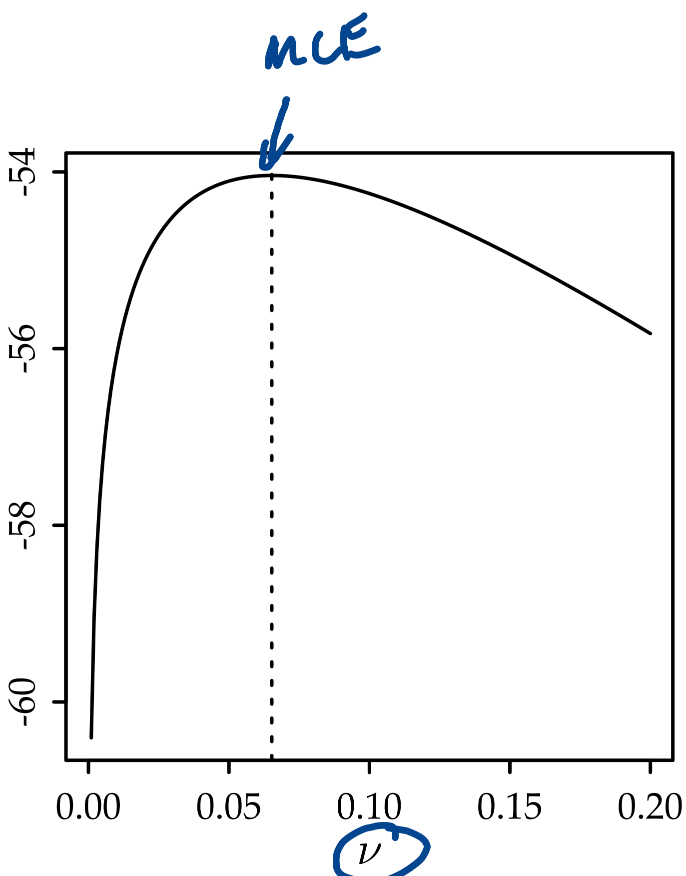
first site: G <--> G

second site: A <--> G

Maximum likelihood estimation

First 32 nucleotides of the ψη-globin gene of gorilla and orangutan:

```
gorilla   GAAGTCCTTGAGAAATAAACTGCACACACTGG
orangutan GGACTCCTTGAGAAATAAACTGCACACACTGG
```



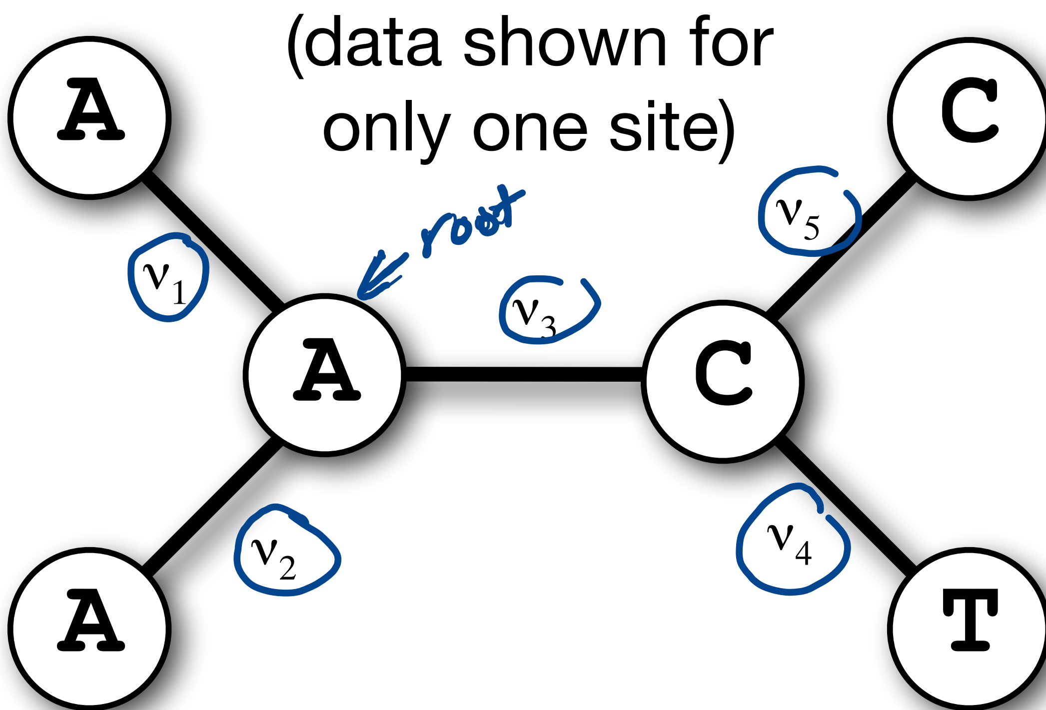
MLE = 0.06525853

Evolutionary distances for several common models

Model	Expected no. substitutions: $v = \{r\}t$
JC69	$\nu = \{3\beta\}t$ $\leftarrow \beta t$
F81	$\nu = \{2\beta\left(\pi_R\pi_Y + \pi_A\pi_G + \pi_C\pi_T\right)\}\, t$ $\leftarrow$
K80	$\nu = \{\beta\left(\kappa + 2\right)\}\, t$ βt given ν, κ
HKY85	$\nu = \{2\beta\left[\pi_R\pi_Y + \kappa\left(\pi_A\pi_G + \pi_C\pi_T\right)\right]\}\, t$

STOPPED HERE FEB. 6, 2024

Likelihood of an unrooted tree

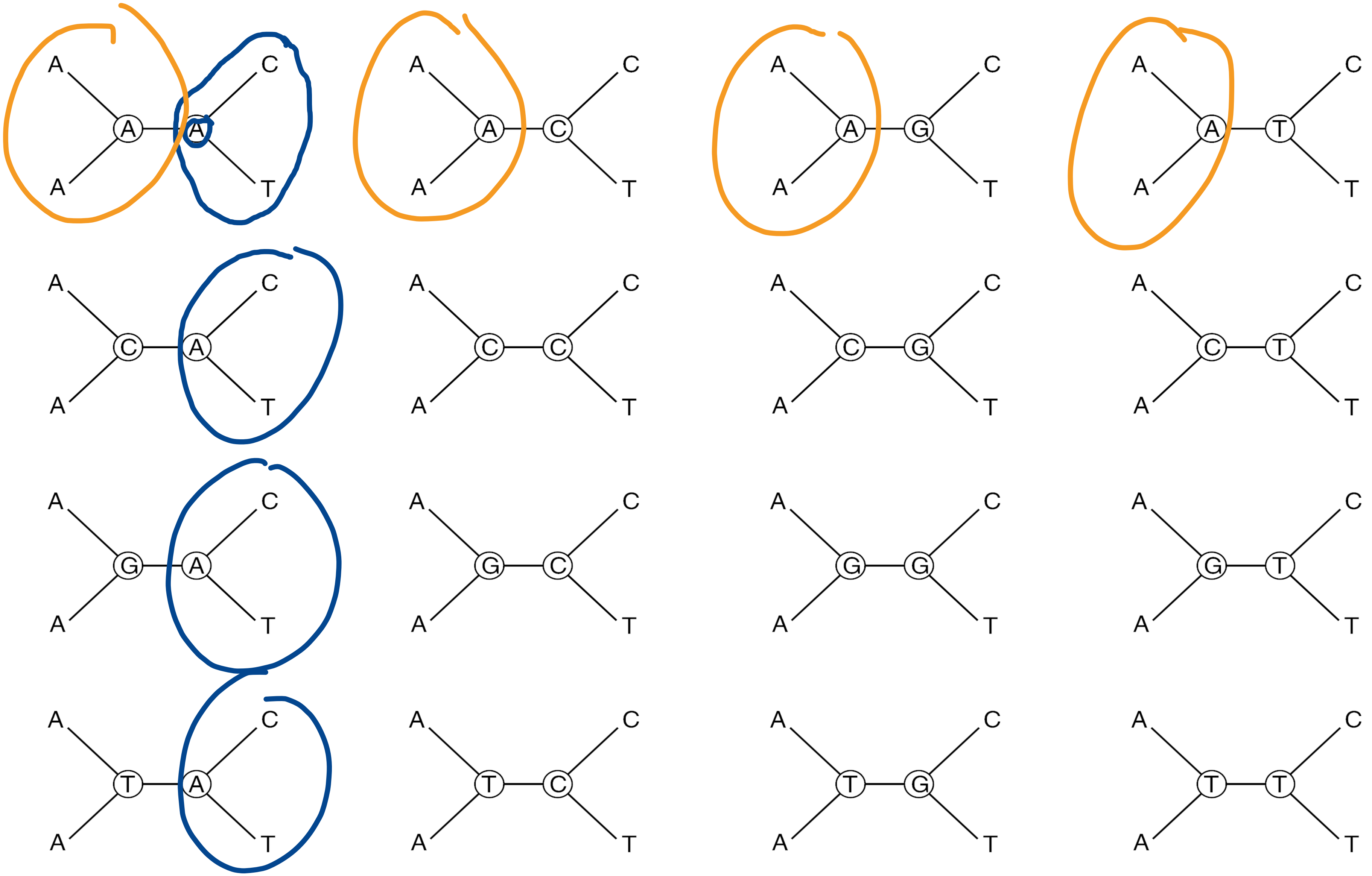


$v_2 = 3\beta_2 t$

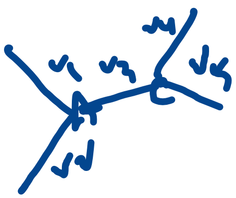
$\beta_2 t = \frac{v_2}{3}$

$$L = \frac{1}{4}\left(\frac{1}{4} + \frac{3}{4}e^{-\frac{4v_2}{3}}\right)\left(\frac{1}{4} + \frac{3}{4}e^{-\frac{4v_1}{3}}\right)\left(\frac{1}{4} - \frac{1}{4}e^{-\frac{4v_3}{3}}\right) \cdot \left(\frac{1}{4} + \frac{3}{4}e^{-\frac{4v_5}{3}}\right)\left(\frac{1}{4} - \frac{1}{4}e^{-\frac{4v_4}{3}}\right)$$

Brute force vs pruning algorithm



Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376



site k

Putting it all together

$L^{(k)} = L^{(k)}_{AA} + L^{(k)}_{AC} + L^{(k)}_{AG} + L^{(k)}_{AT} \cdots L^{(k)}_{TT}$ ← kth site

$\log L^{(k)} = \log\left(L^{(k)}_{AA} + L^{(k)}_{AC} \cdots L^{(k)}_{TT}\right)$

$\log L = \sum_{k=1}^{n} \log L^{(k)} \rightarrow \max L$ maximized over branch lengths

$\prod_{k=1}^{n} L^{(k)}$

$L = L^{(1)} L^{(2)} L^{(3)} \cdots L^{(n)}$

$\log L = \log L^{(1)} + \log L^{(2)} + \cdots + \log L^{(n)}$