

# Review Questions

What is a Markov chain? What is the Markov property?

What are the differences between discrete- and continuous-time Markov chains?

What is the state space for a typical phylogenetic model?

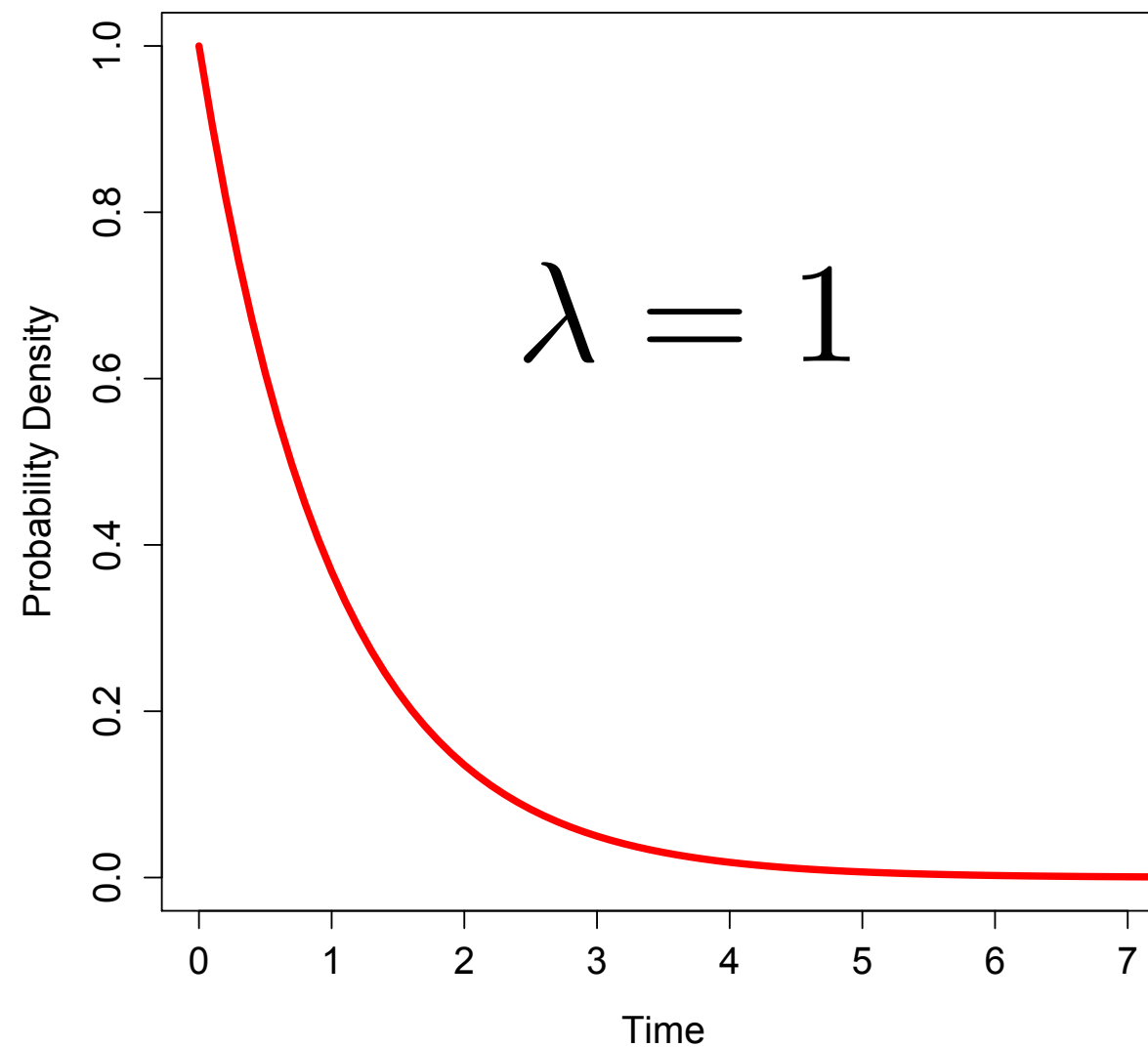
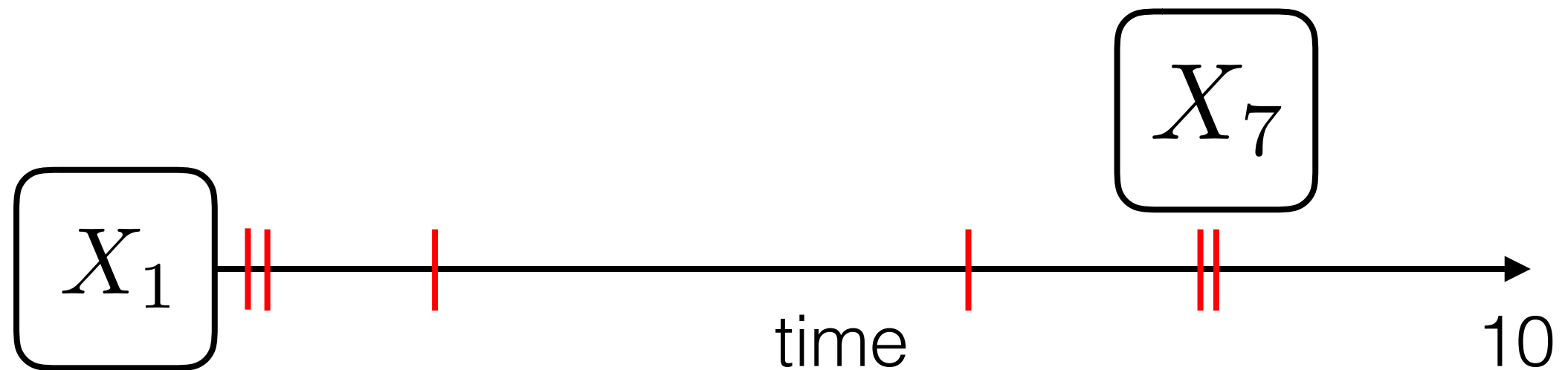
What does the stationary distribution tell us?

What does it mean for a chain to be reversible?

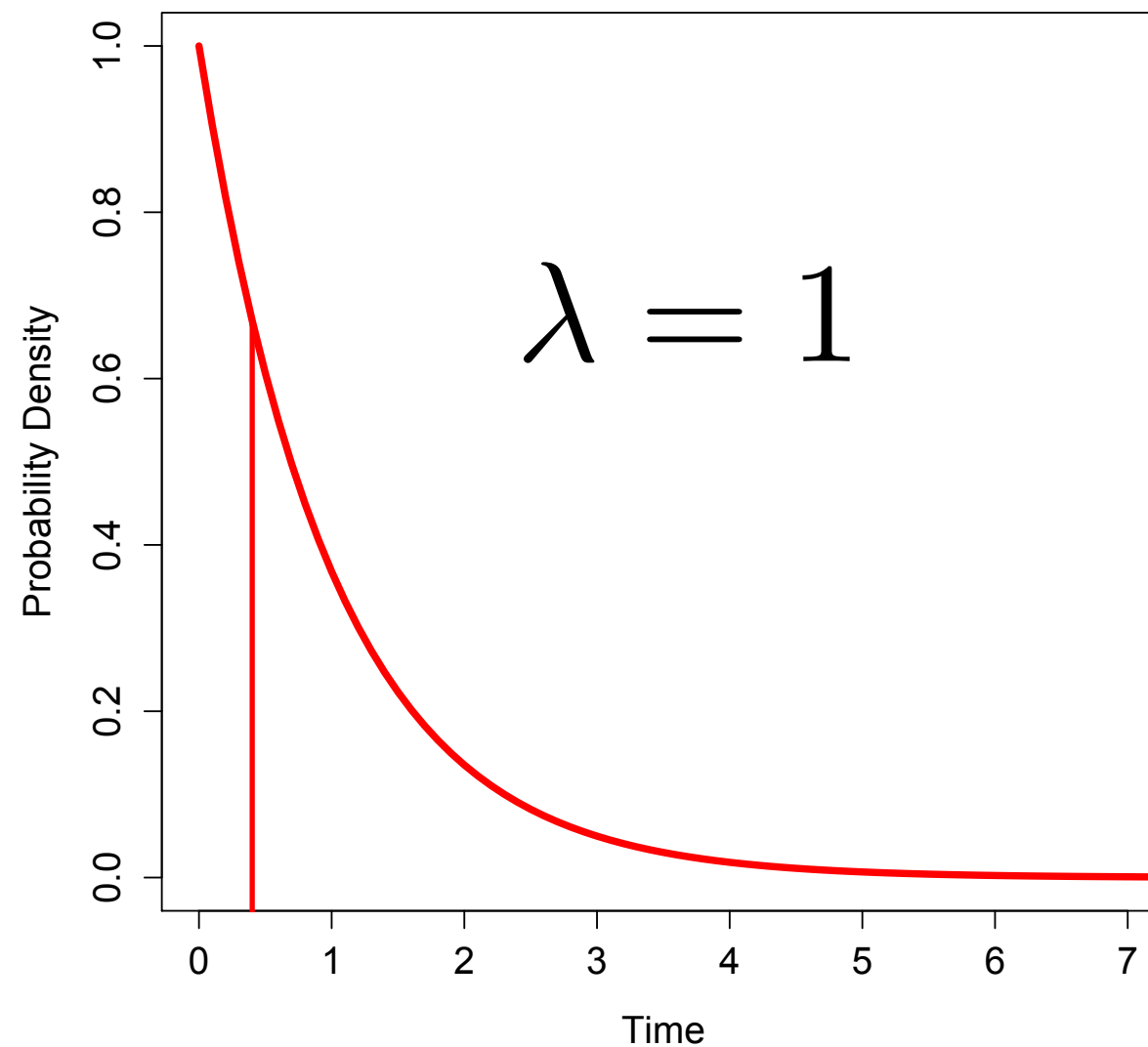
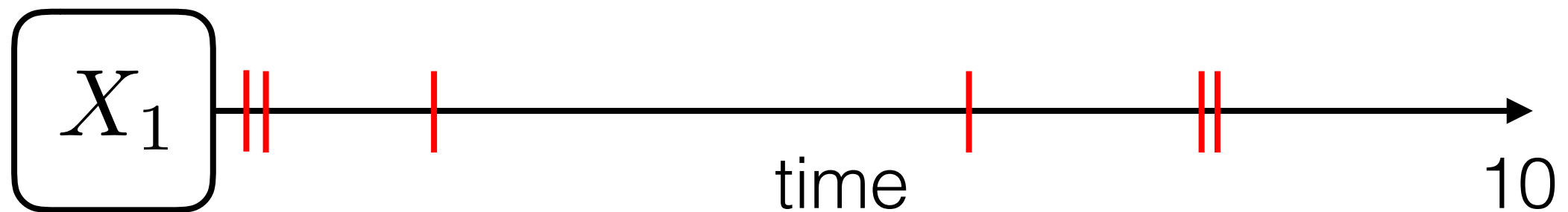
# Simulating Sequence Evolution Assignment

PSEUDOCODE! What are the steps?

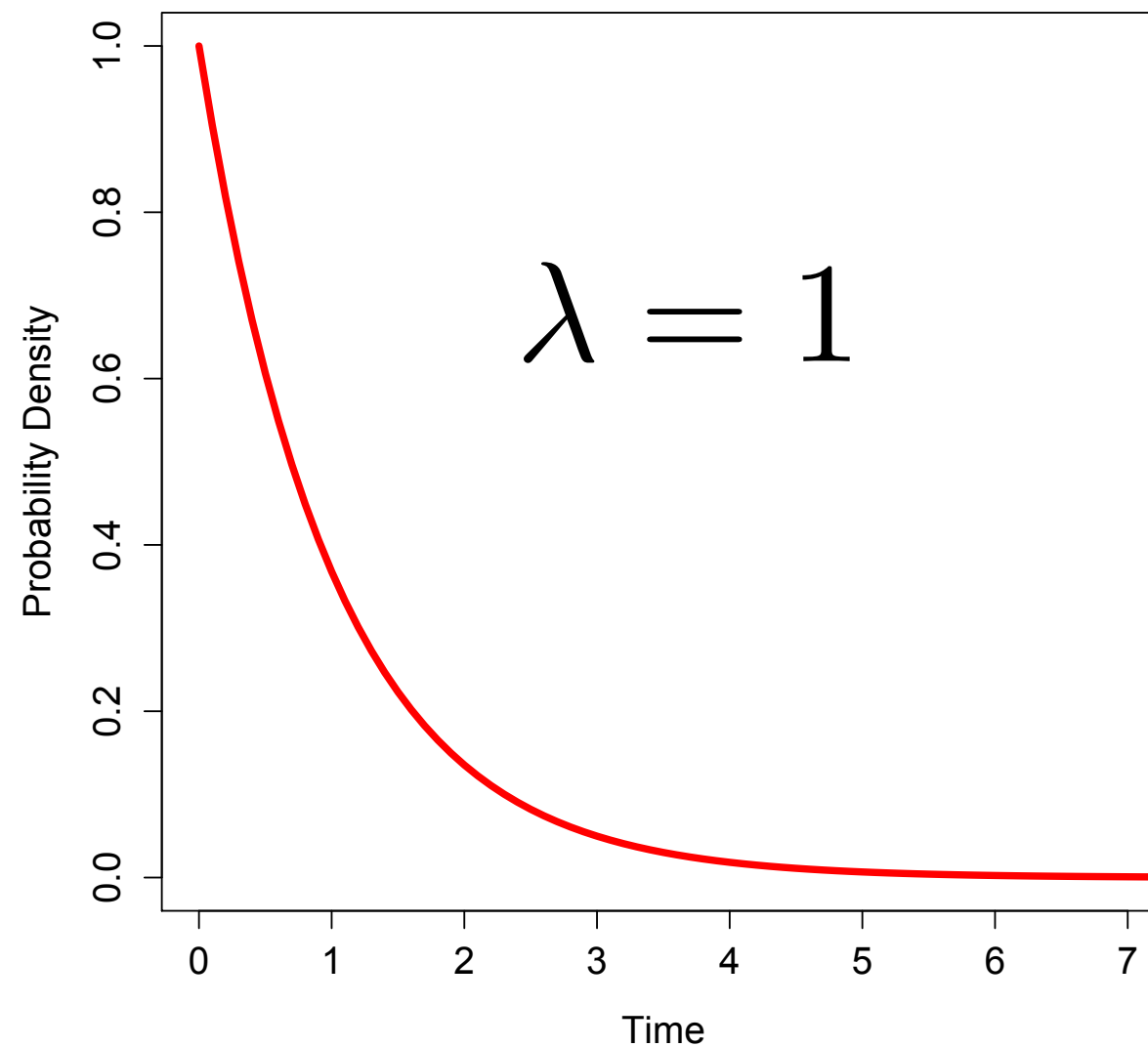
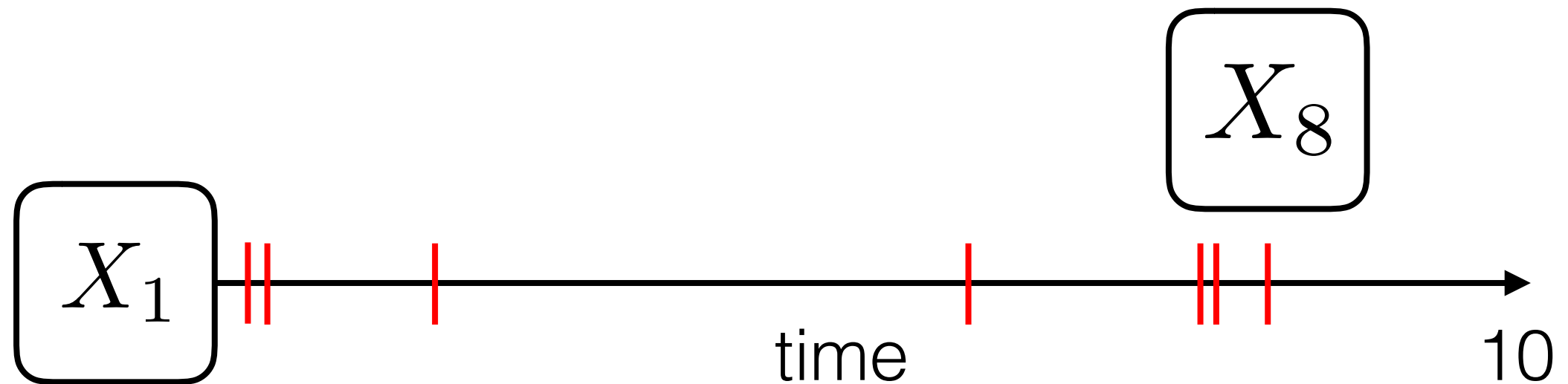
# Continuous-Time Markov Chains



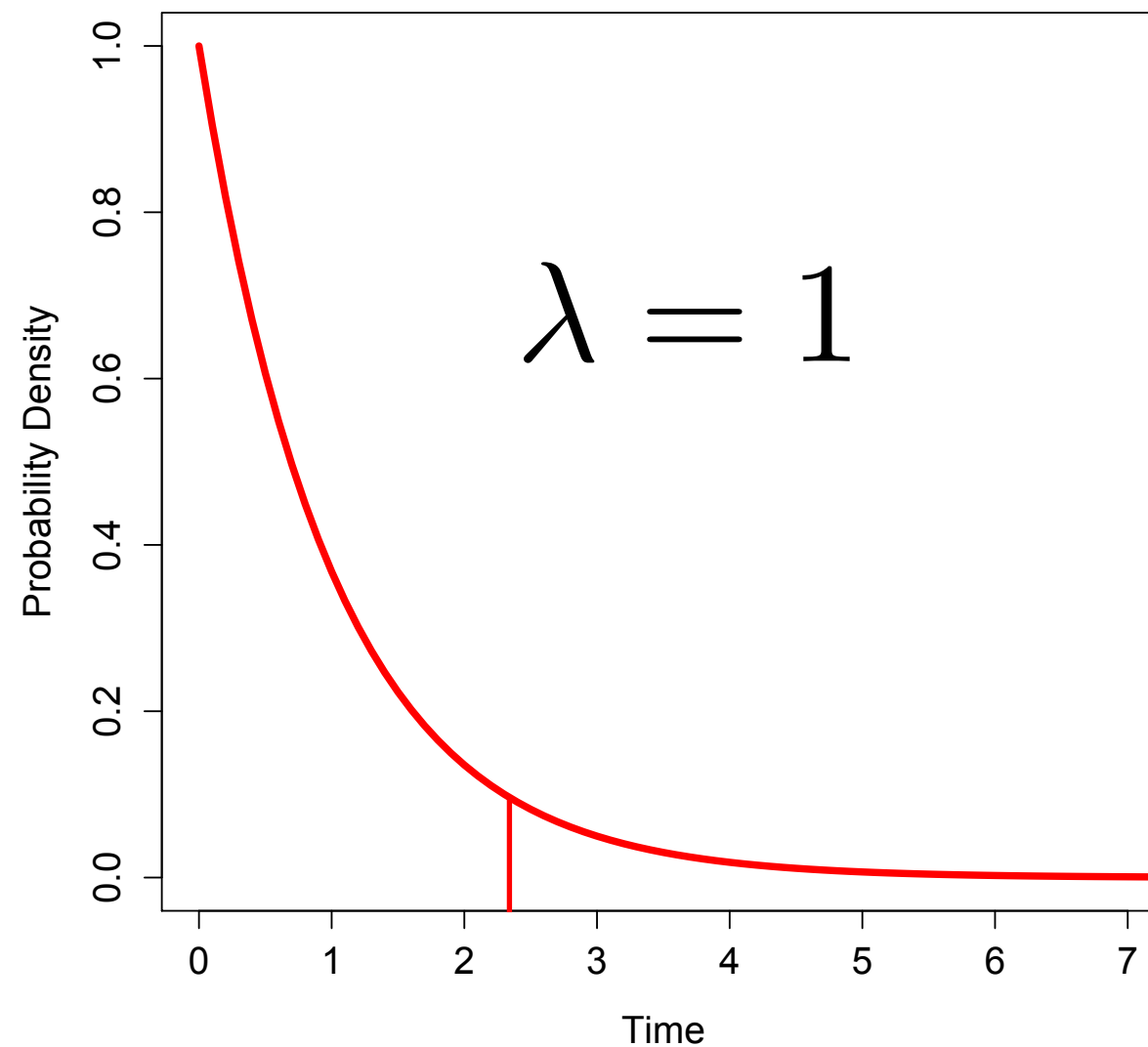
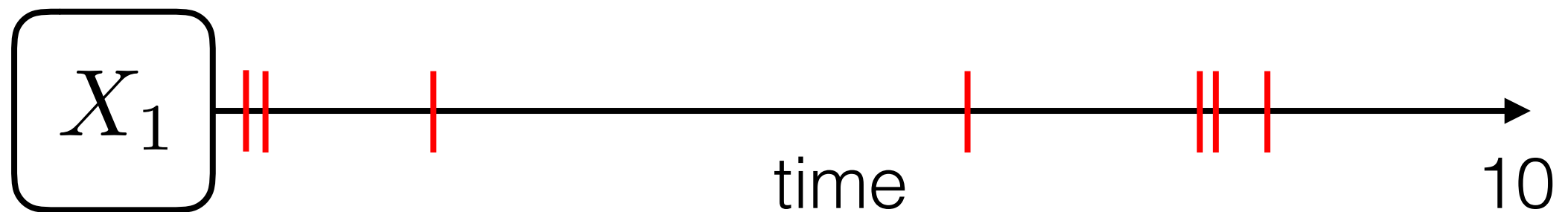
# Continuous-Time Markov Chains



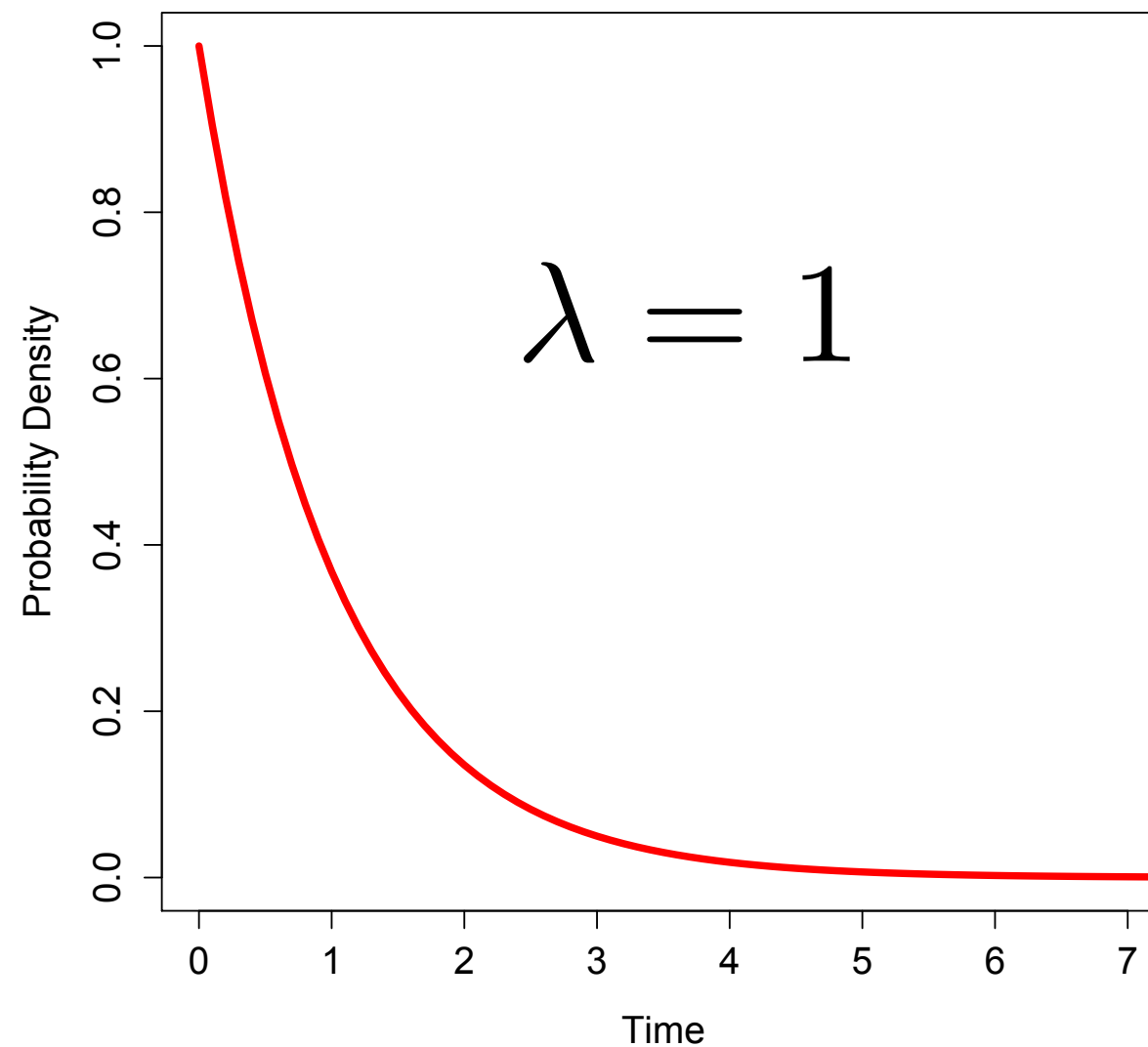
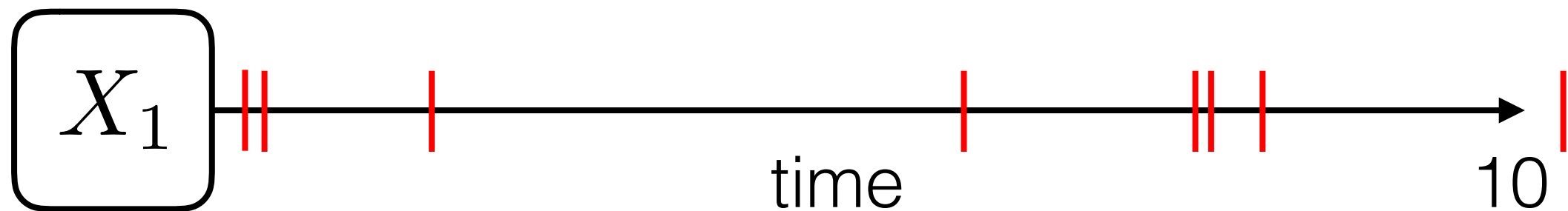
# Continuous-Time Markov Chains



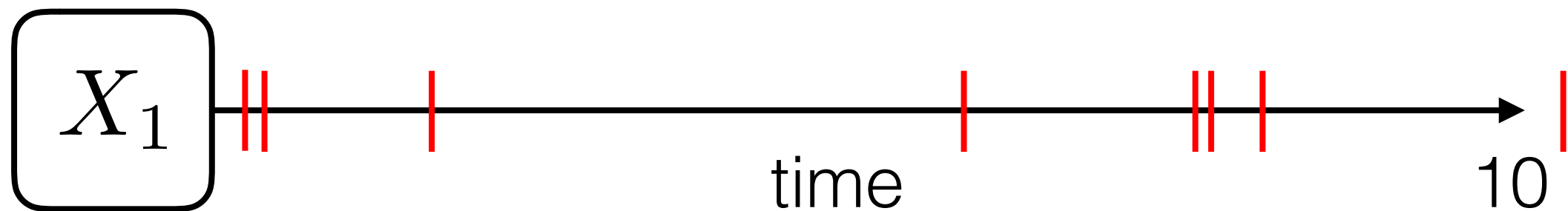
# Continuous-Time Markov Chains



# Continuous-Time Markov Chains



# Continuous-Time Markov Chains

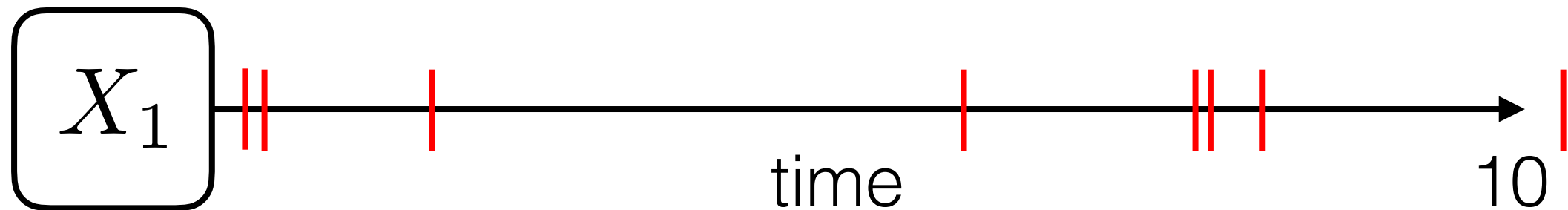


Simulation stops when the total waiting time exceeds the length of the simulation (for phylogenetics, think of this as a branch length).

The state of the chain at the end of the simulation is the state last sampled before the branch length was exceeded ( $X_8$ , in this case).

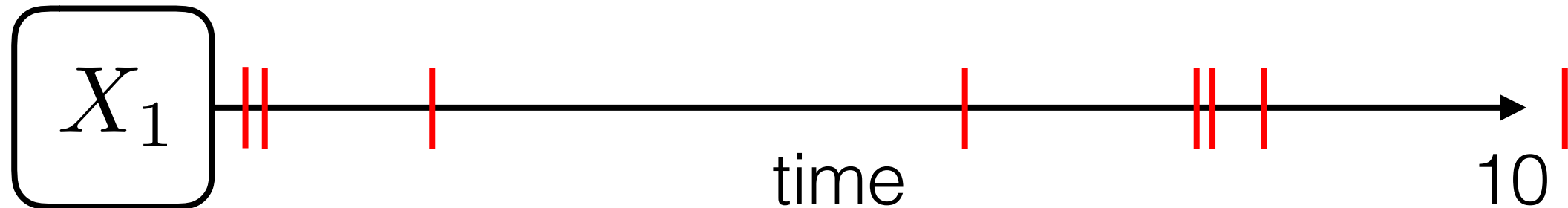


# Continuous-Time Markov Chains



If the rate parameter ( $\lambda$ ) is 1, what is the expected number of events in a Poisson distribution? What does this mean if we're thinking about a phylogenetic tree with branch lengths?

# Poisson Process



A continuous-time Markov chain is one type of a **Poisson process**, which has these properties:

- (1) The number of events that occur in an interval of length  $t$  is a  $\text{Pois}(\lambda t)$  random variable.
- (2) The number of events that occur in disjoint intervals are independent of one another.
- (3) The waiting times between events are distributed as:  $T \sim \text{Exp}(\lambda)$

# CTMC Rate Matrix

$$\begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix}$$

Jukes and Cantor (1969)


# CTMC Rate Matrix

$$\begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix}$$

Jukes and Cantor (1969)

# CTMC Rate Matrix

Exponential Rate  
for State A


$$\begin{matrix} & A & C & G & T \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix} \end{matrix}$$

Jukes and Cantor (1969)

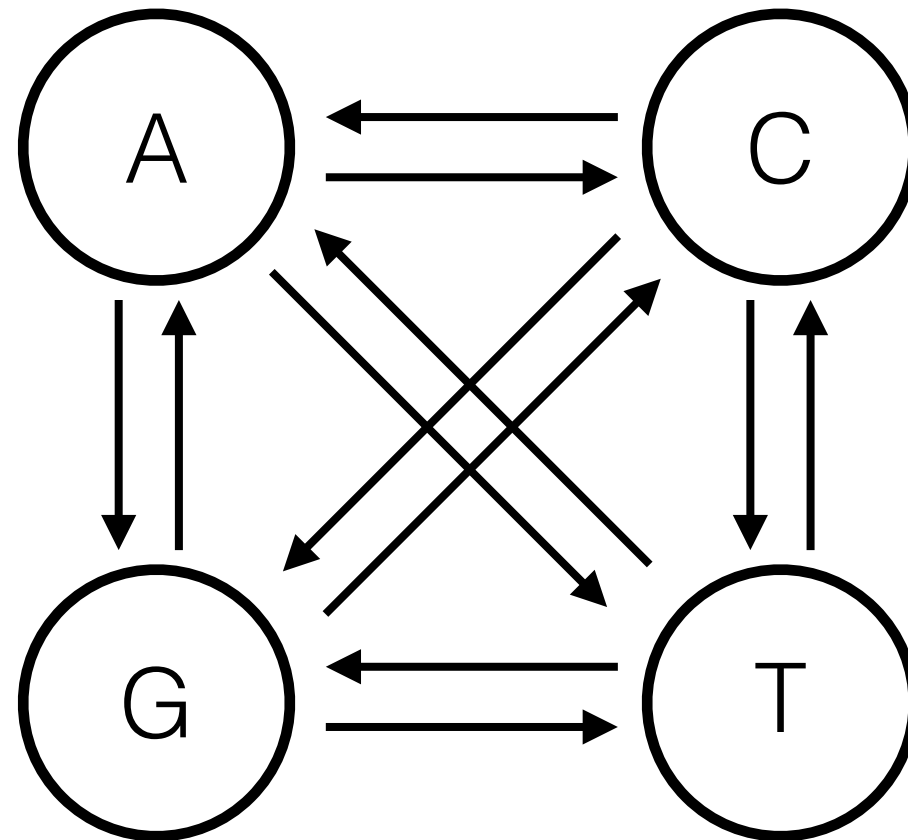
# CTMC Rate Matrix

Relative  
Probabilities of  
Transition

$$\begin{matrix} & A & C & G & T \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/3 & -1 & 1/3 & 1/3 \\ 1/3 & 1/3 & -1 & 1/3 \\ 1/3 & 1/3 & 1/3 & -1 \end{pmatrix} \end{matrix}$$

Jukes and Cantor (1969)

# Jukes and Cantor (1969)



# CTMC Rate Matrix

$$\begin{array}{c}
 A \\
 C \\
 G \\
 T
 \end{array}
 \begin{array}{c}
 A \qquad C \qquad G \qquad T \\
 \left( \begin{array}{cccc}
 & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\
 \pi_A r_{AC} & & \pi_G r_{CG} & \pi_T r_{CT} \\
 \pi_A r_{AG} & \pi_C r_{CG} & & \pi_T r_{GT} \\
 \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} & 
 \end{array} \right)
 \end{array}$$

General Time Reversible (GTR; Tavaré 1986)



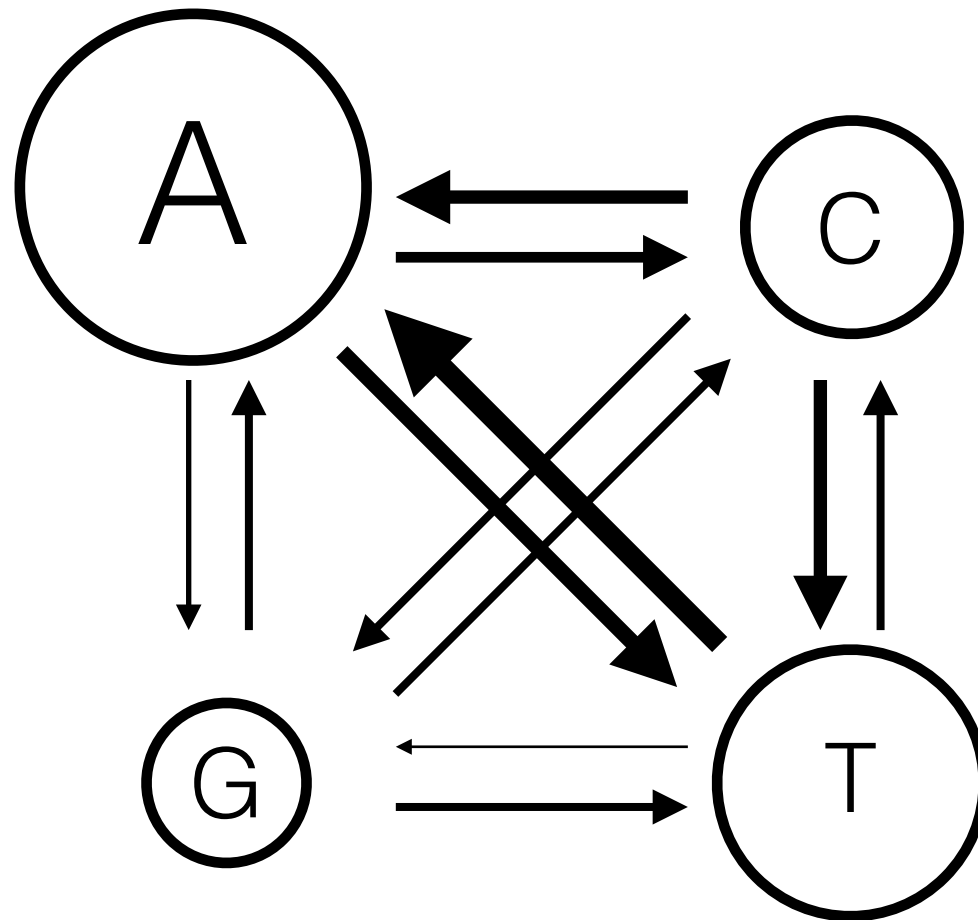
# CTMC Rate Matrix

$$\pi = (\pi_A, \pi_C, \pi_G, \pi_T) \quad R = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} & r_{AC} & r_{AG} & r_{AT} \\ & & r_{CG} & r_{CT} \\ & & & r_{GT} \end{pmatrix} \end{matrix}$$

$$Q = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & & & \\ \pi_A r_{AG} & \pi_C r_{CG} & & \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} & \end{pmatrix} \end{matrix}$$

General Time Reversible (GTR; Tavaré 1986)

# General Time Reversible (GTR)



# Reversibility

$$\pi_i q_{ij} = \pi_j q_{ji}, \text{ for all } i \neq j$$

$$Q = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & & \pi_G r_{CG} & \pi_T r_{CT} \\ \pi_A r_{AG} & \pi_C r_{CG} & & \pi_T r_{GT} \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} & \end{pmatrix} \end{matrix}$$

if  $i = A$  and  $j = C$ ,

$$\pi_A q_{AC} = \pi_C q_{CA}$$

$$\pi_A \pi_C r_{AC} = \pi_C \pi_A r_{AC}$$

# Branch-Length Scaling

$$\begin{array}{c}
 A \\
 C \\
 G \\
 T
 \end{array}
 \begin{pmatrix}
 & A & C & G & T \\
 & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\
 \pi_A r_{AC} & & \pi_G r_{CG} & \pi_T r_{CT} \\
 \pi_A r_{AG} & \pi_C r_{CG} & & \pi_T r_{GT} \\
 \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} &
 \end{pmatrix}$$

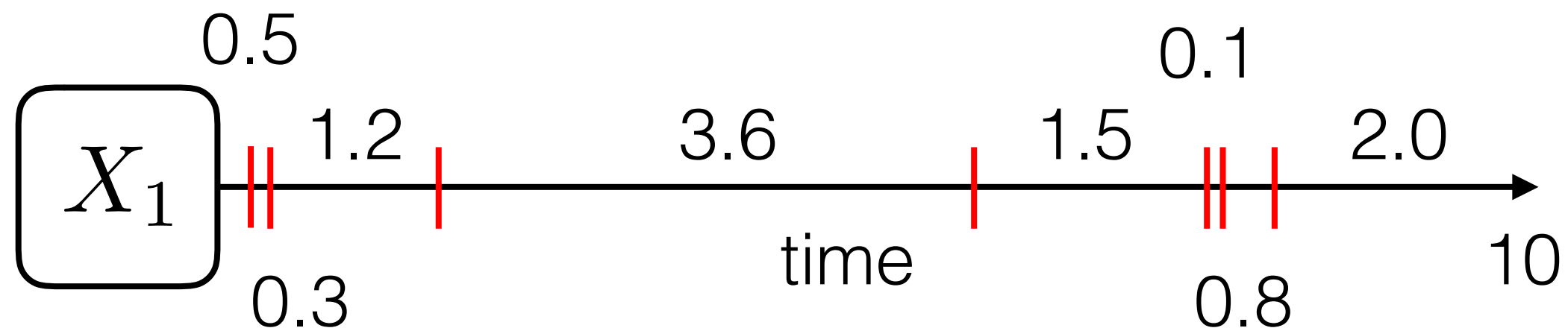
*Must average to 1*

Branch lengths typically denote **expected number of substitutions**. For this to be true, the weighted mean across all states must be 1.

# Probability of a Character History

Jukes-Cantor

Instead of thinking about simulating a character history, let's calculate the probability of one that we've been given.

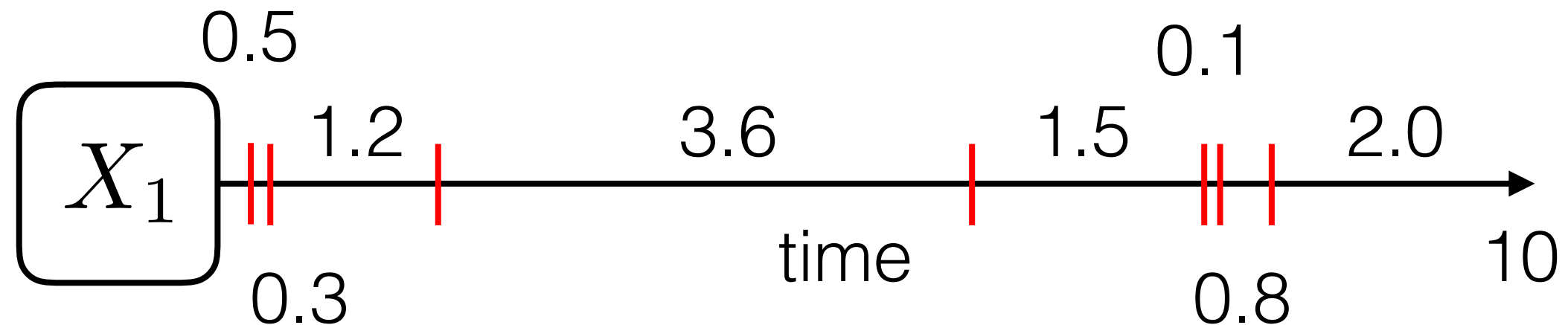


The events in this history are independent. When events are independent how do we combine their probabilities?

What are the relevant probabilities here?

# Probability of a Character History

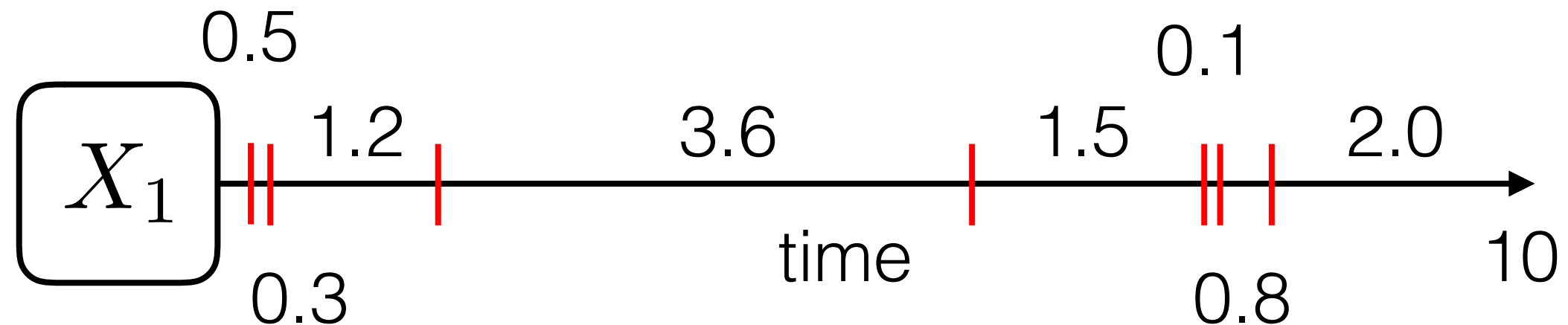
Jukes-Cantor



$$\text{Prob}(\text{history}) = \text{Exp}(0.5)\text{Exp}(0.3)\text{Exp}(1.2)\text{Exp}(3.6)\text{Exp}(1.5)\text{Exp}(0.1)\text{Exp}(0.8)e^{-2.0}(1/3)^7$$

# Probability of a Character History

Jukes-Cantor

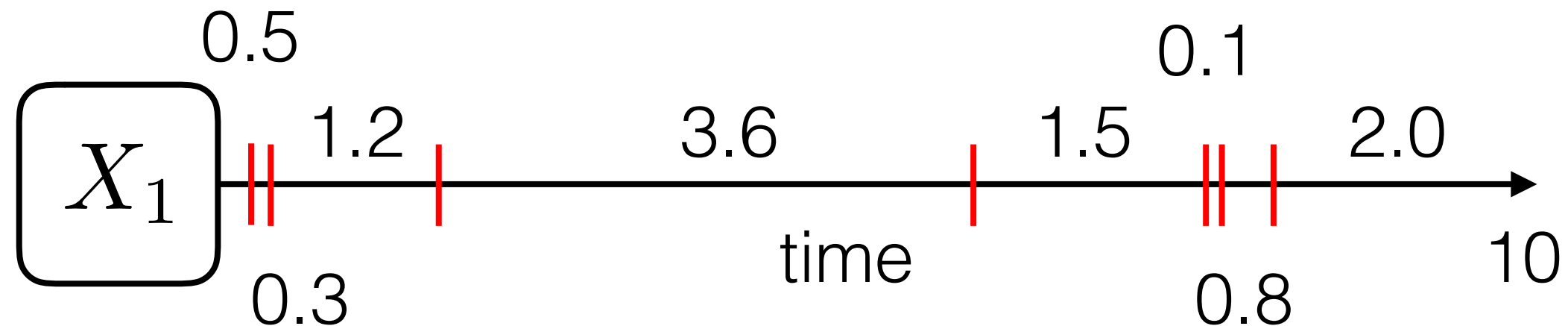


$$\text{Prob}(\text{history}) = \text{Exp}(0.5)\text{Exp}(0.3)\text{Exp}(1.2)\text{Exp}(3.6)\text{Exp}(1.5)\text{Exp}(0.1)\text{Exp}(0.8)e^{-2.0}(1/3)^7$$

Waiting Times

# Probability of a Character History

Jukes-Cantor



$$\text{Prob}(\text{history}) = \text{Exp}(0.5)\text{Exp}(0.3)\text{Exp}(1.2)\text{Exp}(3.6)\text{Exp}(1.5)\text{Exp}(0.1)\text{Exp}(0.8)e^{-2.0}(1/3)^7$$

Waiting Times

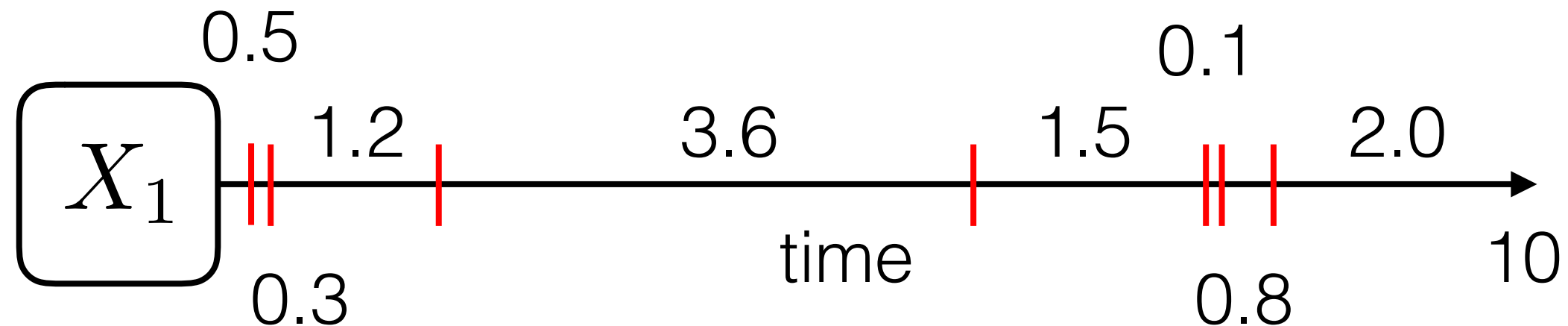
Prob No  
Change  
at End

Integral of  $\text{Exp}(1)$  from 2.0 to infinity



# Probability of a Character History

Jukes-Cantor



$$\text{Prob}(\text{history}) = \text{Exp}(0.5)\text{Exp}(0.3)\text{Exp}(1.2)\text{Exp}(3.6)\text{Exp}(1.5)\text{Exp}(0.1)\text{Exp}(0.8)e^{-2.0}(1/3)^7$$

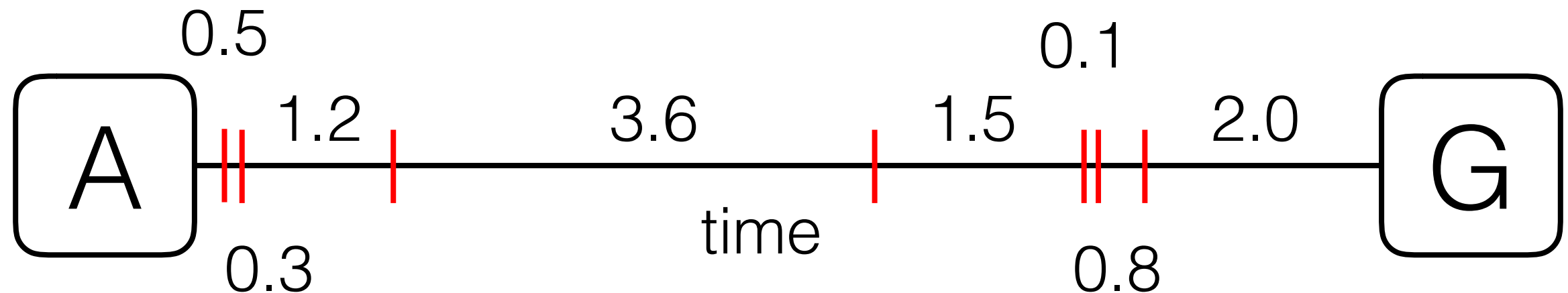
Waiting Times

Prob No  
Change  
at End

Character  
State  
Changes

# Probability of Starting and Ending

Jukes-Cantor



Let's say we simulated along our branch of length 10. We started in A and ended in G, with those specific events marked.

# Probability of Starting and Ending

Jukes-Cantor



Let's say we simulated along our branch of length 10. We started in A and ended in G, with those specific events marked.

But now let's say we don't actually know the precise character history. How do we calculate

$$P(A \text{ to } G | t = 10) ?$$

# Probability of Starting and Ending

Jukes-Cantor



$$P(t) = e^{Qt}$$

This is the continuous-time equivalent of the matrix exponentiation we did with the discrete-time matrix!