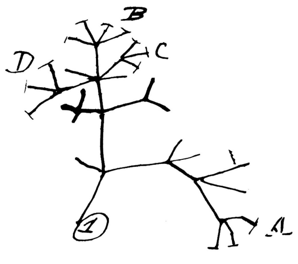


IPS-164 INTRODUCTION TO PHYLOGENETICS 2022

Lecture 7
Intro to statistical phylogenetics. Part II

Sergei Tarasov

Beetle curator & Docent

Finnish Museum of Natural History, University of Helsinki

- @tarasov_sergio
- sergei.tarasov@helsinki.fi
- https://www.tarasovlab.com

PLAN OF THE TODAY'S LECTURE

Aim: Derive a Markov model of trait (DNA) evolution and its likelihood inference on phylogenetic tree

1. From Binomial distribution to Markov model
2. We will study general principles of Markov models
3. We will study Markov models on phylogenetic tree; Felsenstein's prunning algorithm

Binomial model (and distribution)

Binomial model gives the probability of seeing *k* heads in *n* coin tosses (trials) given that probability of seeing a head in one coin toss is *p*.

Number of ways to choose *k* heads

Probability of seeing *k* heads

Number of tosses in an experiment

$$B(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Number of heads in an experiment

Probability of seeing heads

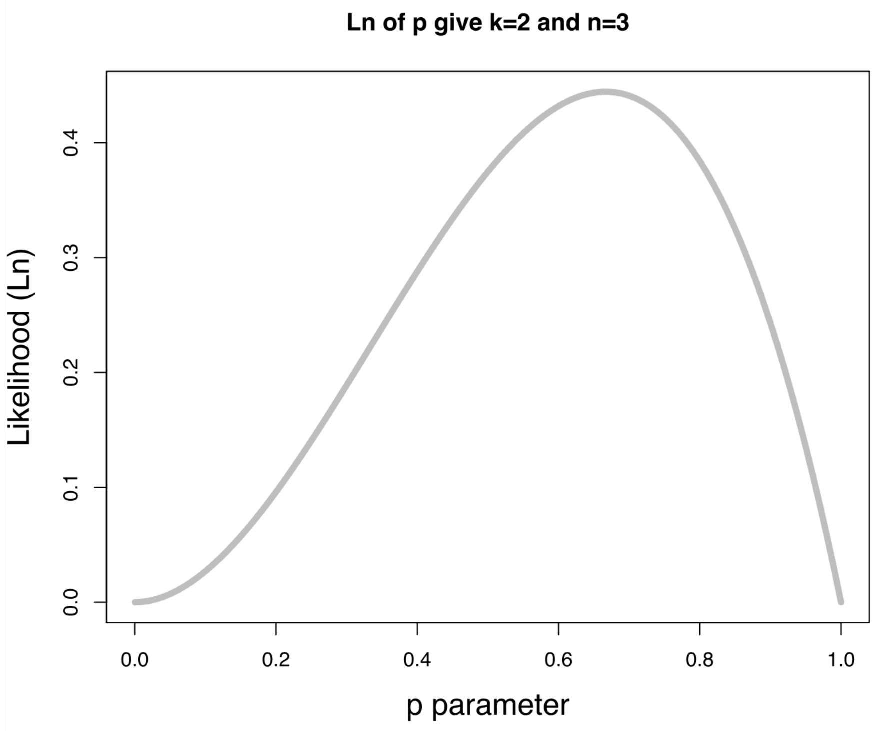
Probability of seeing tails (*n*-*k*)

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

- Coin is fair
- We toss the coin 3 times

Likelihood function of Binomial distribution

In statistics, a likelihood function (often simply the likelihood) is a particular function of the parameter of a statistical model given data. Likelihood functions play a key role in statistical inference.



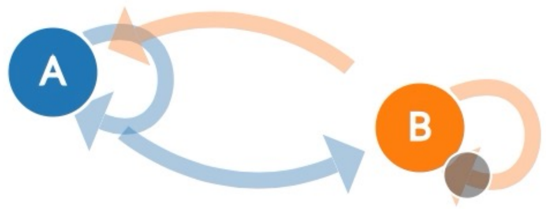
Binomial Likelihood:
Given *n* and *k* infer *p* that maximizes the likelihood function

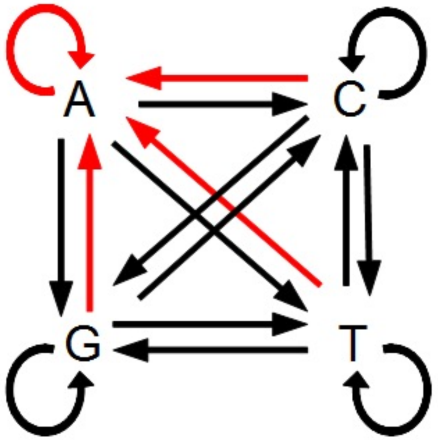
$$Ln(p \mid n=3, k=2) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{3}{2} p^2 (1-p)^{3-2}$$

- Domain of *p* is a value between 0 and 1 (since *p* is a probability)
- Let's try all *p*'s to get a likelihood function
- Likelihood function is not a distribution

Modeling phylogenetic process: Markov models (Markov chains)

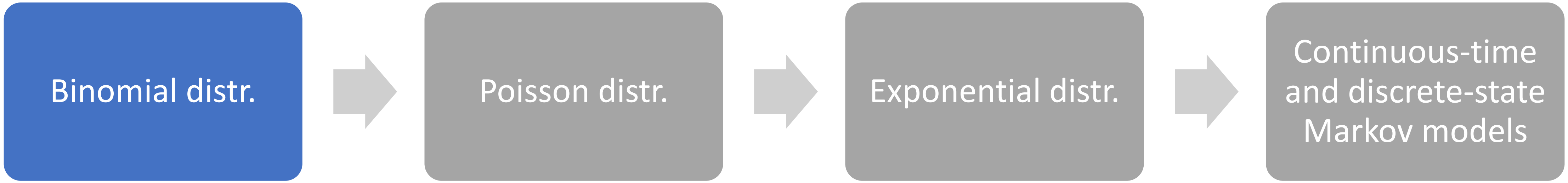
In probability theory, a Markov model is a stochastic model used to model a system that randomly changes from one state to another over time



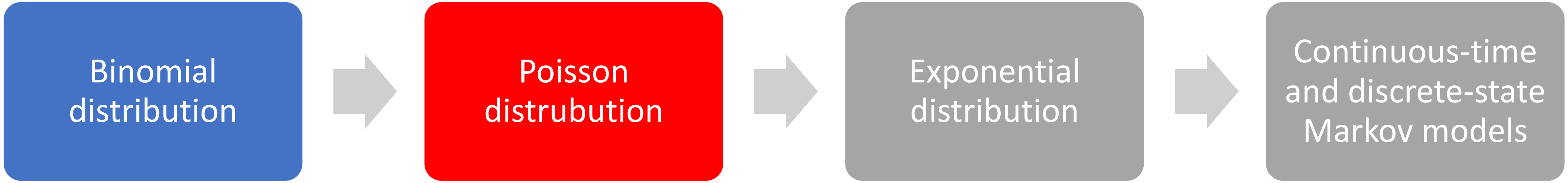


http://setosa.io/ev/markov-chains/

Ingredients to derive continuous-time Markov models

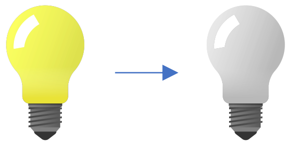


Ingredients to derive continuous-time Markov models



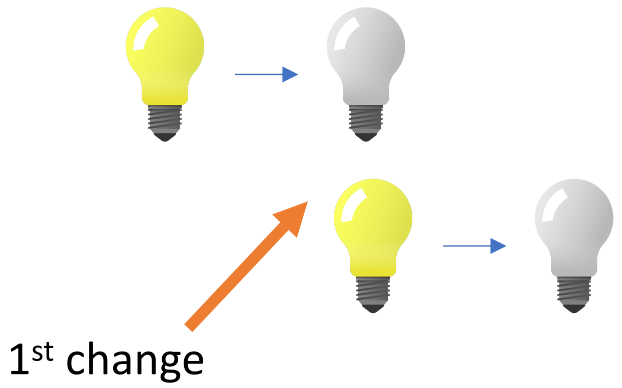
Let’s derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- ‘Bulb experiment’ is a good example for deriving this distribution



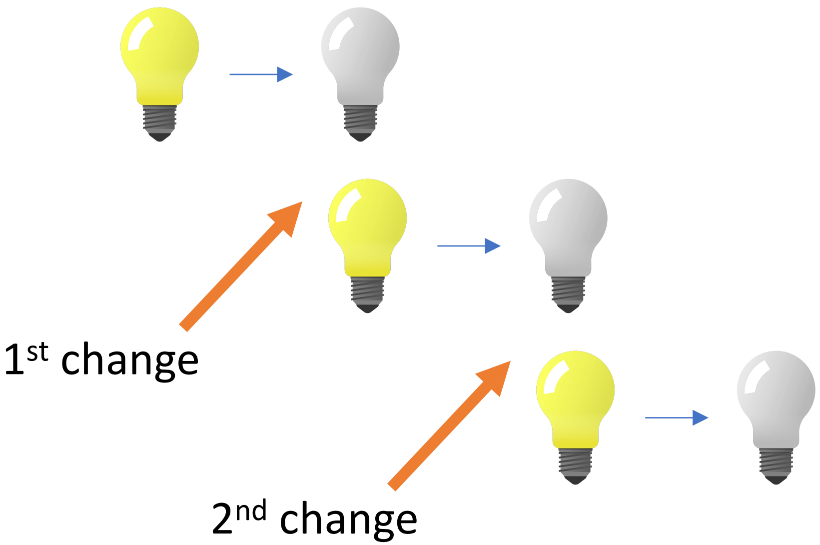
Let’s derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- ‘Bulb experiment’ is a good example for deriving this distribution



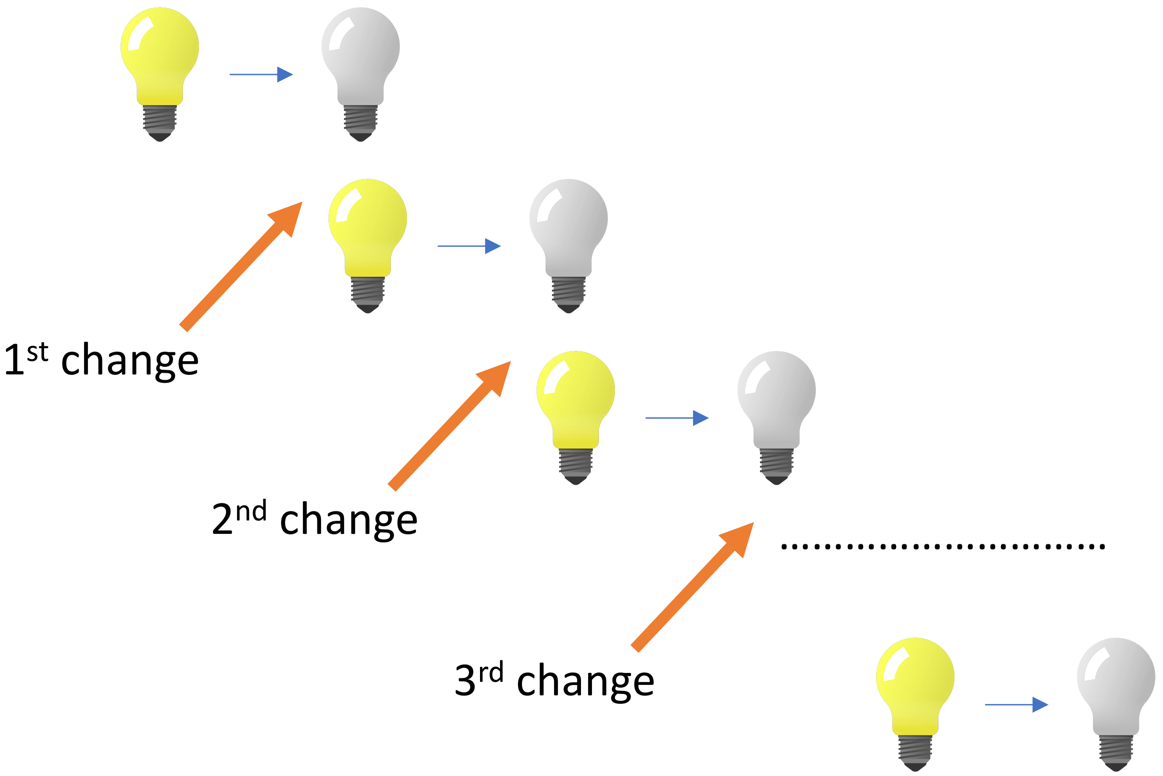
Let's derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- 'Bulb experiment' is a good example for deriving this distribution



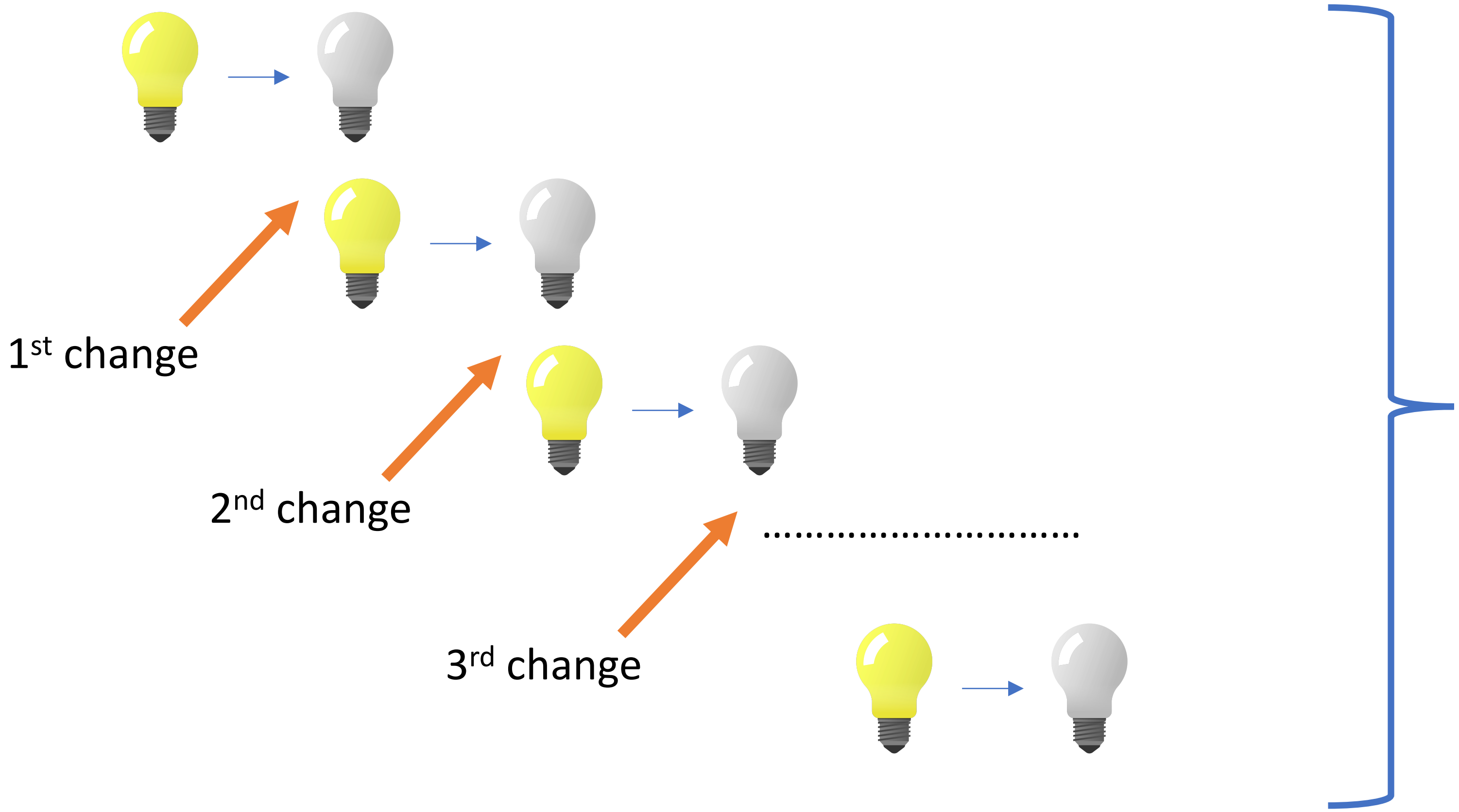
Let’s derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- ‘Bulb experiment’ is a good example for deriving this distribution

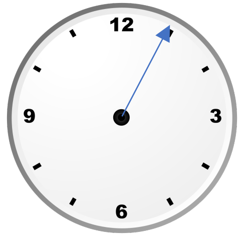


Let’s derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- ‘Bulb experiment’ is a good example for deriving this distribution

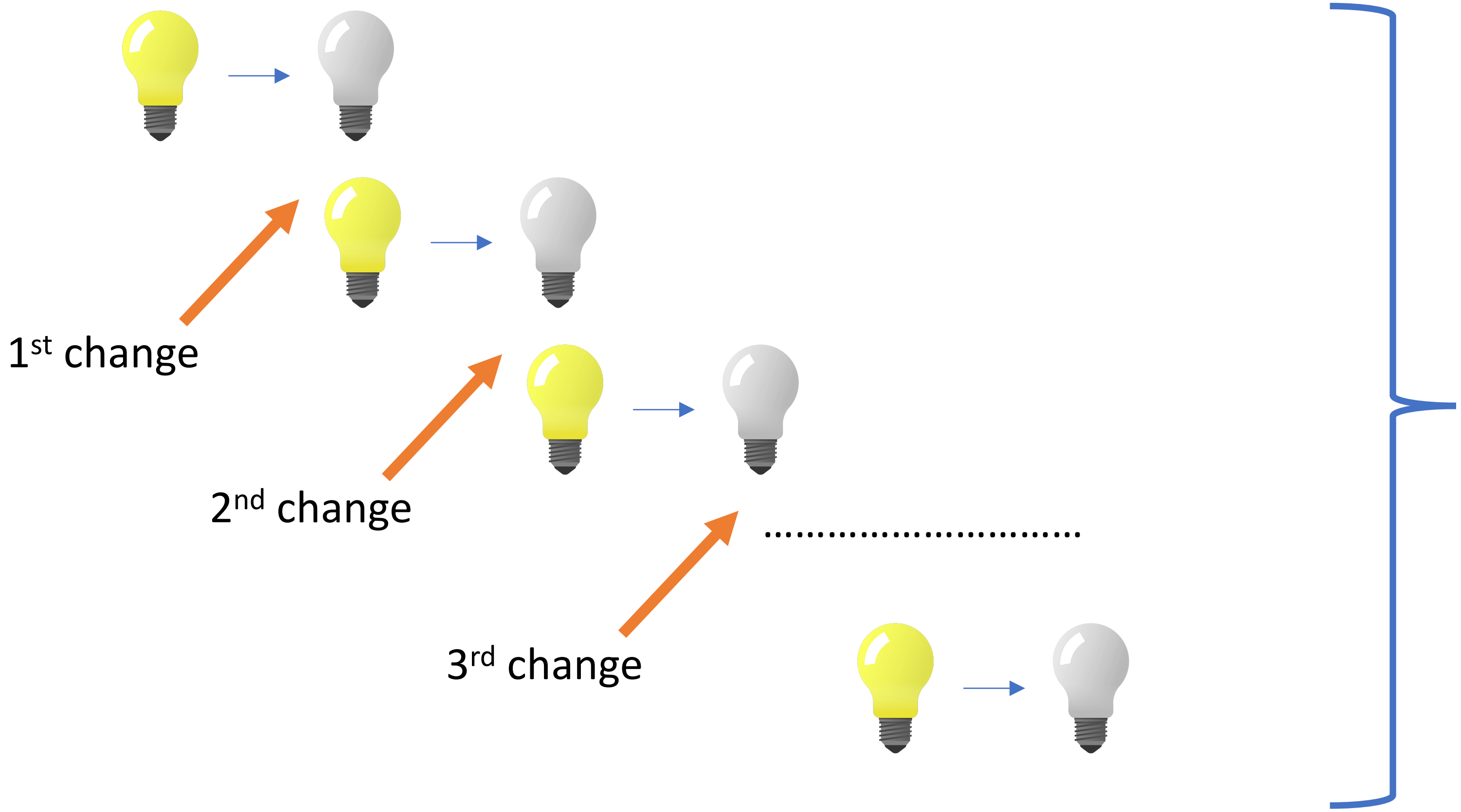


- Fixed amount of time.
- For example 1 year.
- We repeat this experiment 1000 times

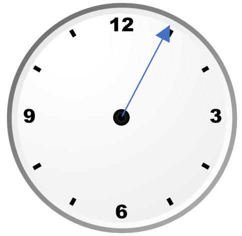


Let's derive a distribution that models number of some evets over time

- We need to translate probabilities into rates
- 'Bulb experiment' is a good example for deriving this distribution

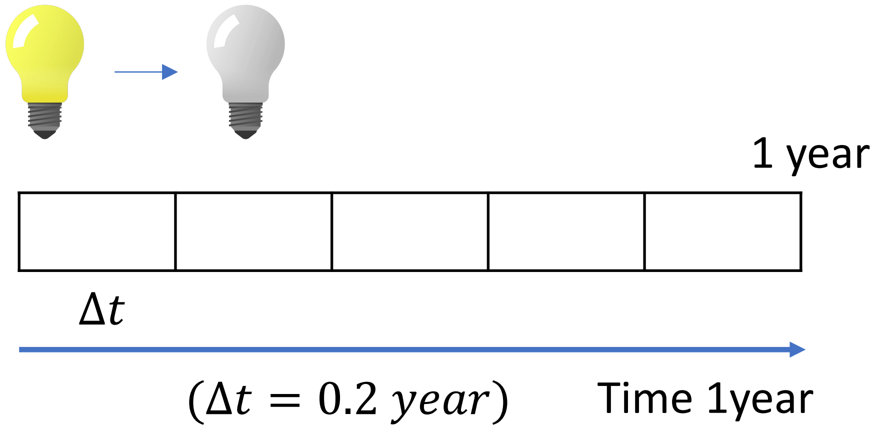


- Fixed amount of time.
- For example 1 year.
- We repeat this experiment 1000 times

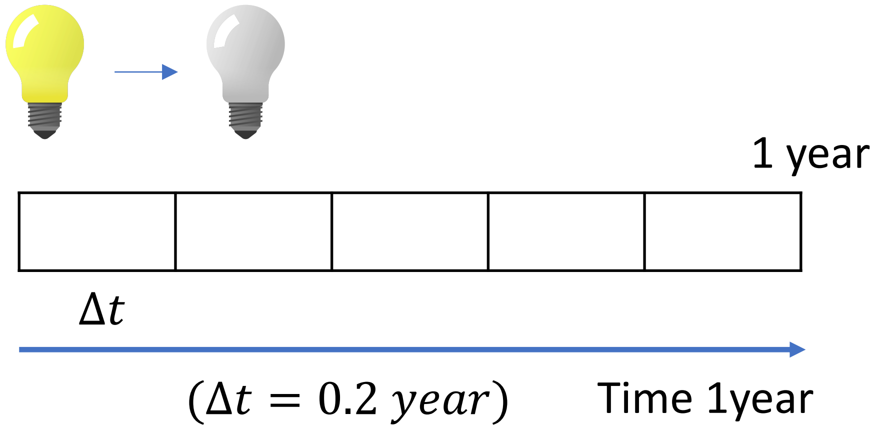


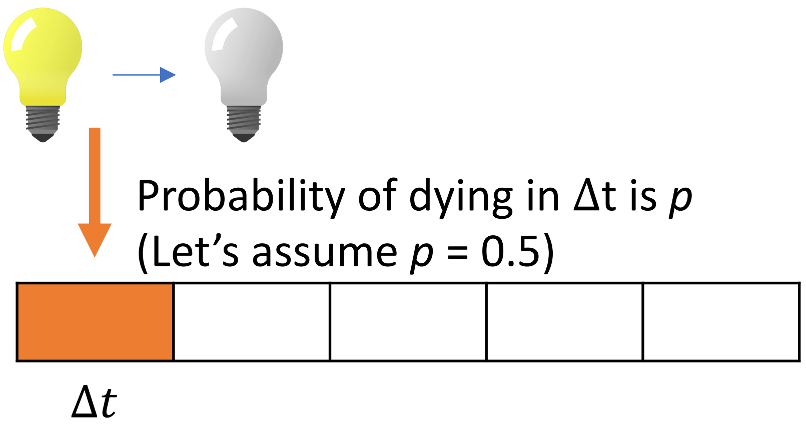
Find a distribution for the number of changes (=number of bulbs that died during one year)!

From Binomial to Poisson distribution

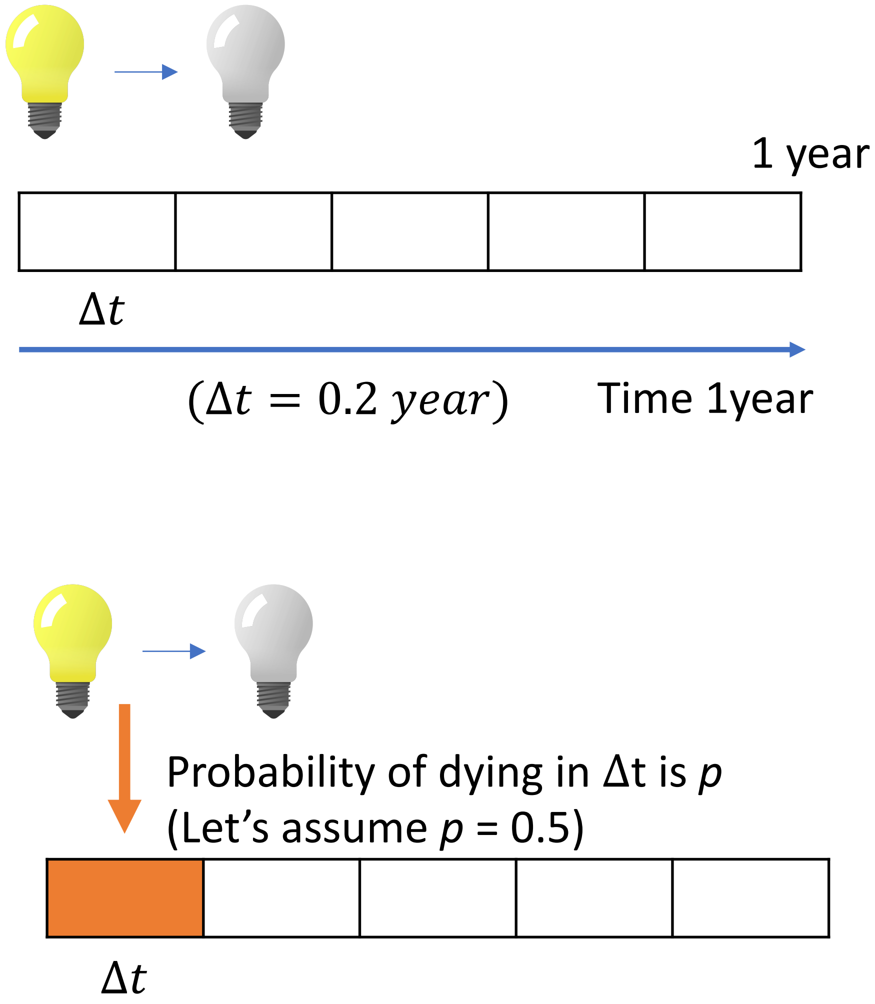


From Binomial to Poisson distribution



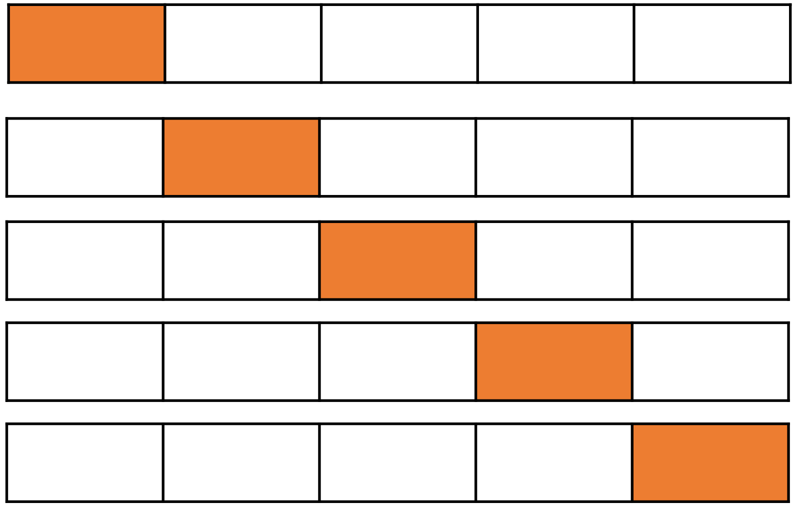


From Binomial to Poisson distribution

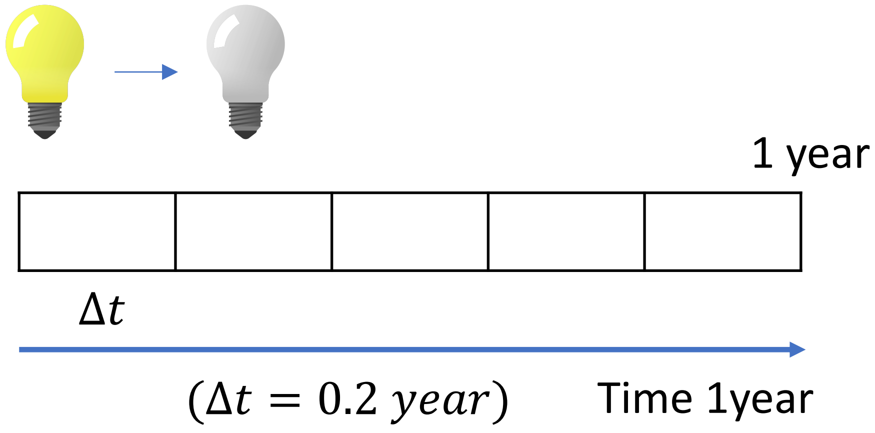


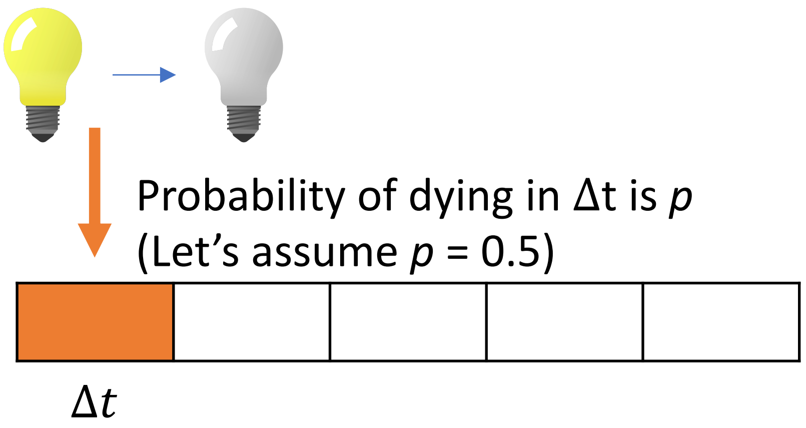
Let's calculate probability of one death (*k*=1)

- One death can happen via 5 different ways
- What distribution is that?



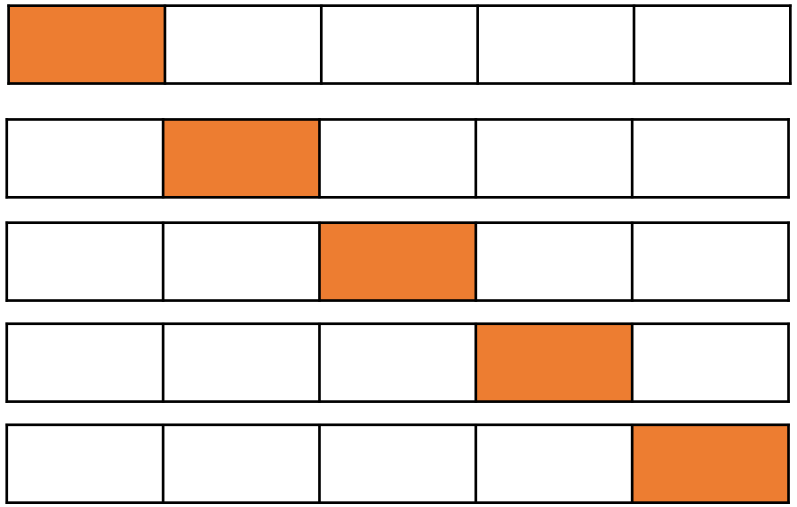
From Binomial to Poisson distribution





Let's calculate probability of one death (*k*=1)

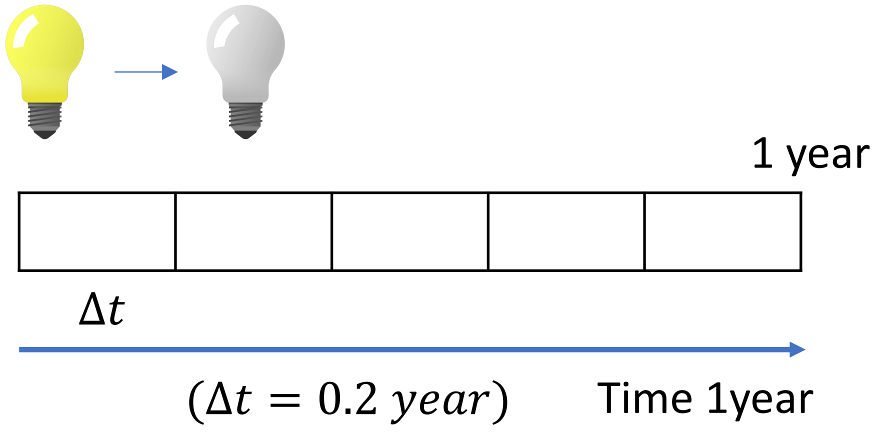
- One death can happen via 5 different ways
- What distribution is that?

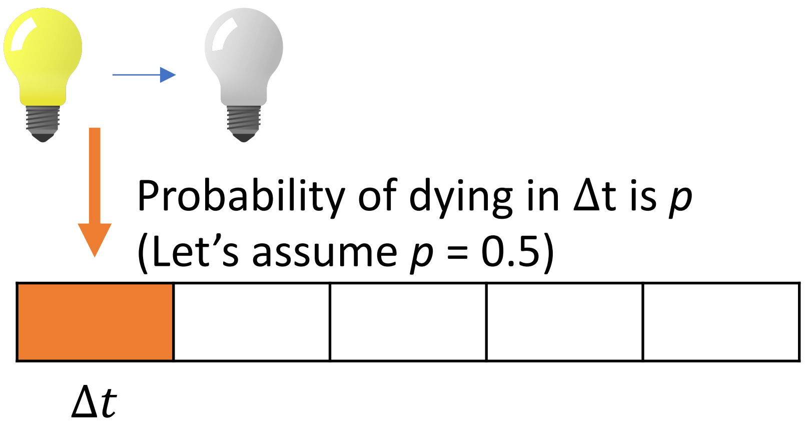


It's Binomial distribution

$$B(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

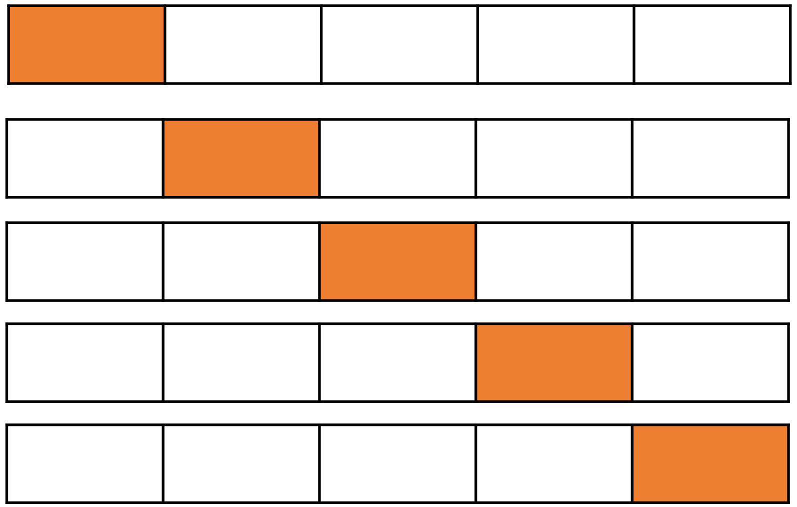
From Binomial to Poisson distribution





Let's calculate probability of one death (*k*=1)

- One death can happen via 5 different ways
- What distribution is that?

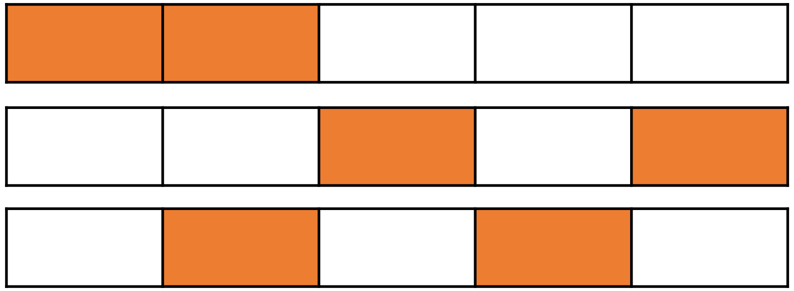


It's Binomial distribution

$$B(k = 1 | n = 5, p = 0.5) = 5 * 0.5^1 0.5^4$$

From Binomial to Poisson distribution

- Probability of two deaths (*k*=2)



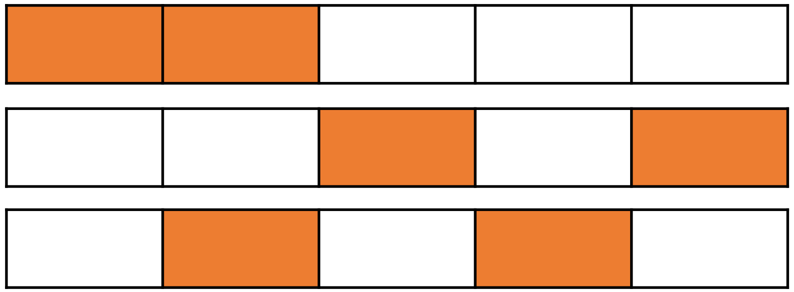
…………………… all other possibilities

It's Binomial distribution

$$B(k = 2|n = 5, p = 0.5) = \binom{5}{2} 0.5^2 0.5^3$$

From Binomial to Poisson distribution

- Probability of two deaths (*k*=2)



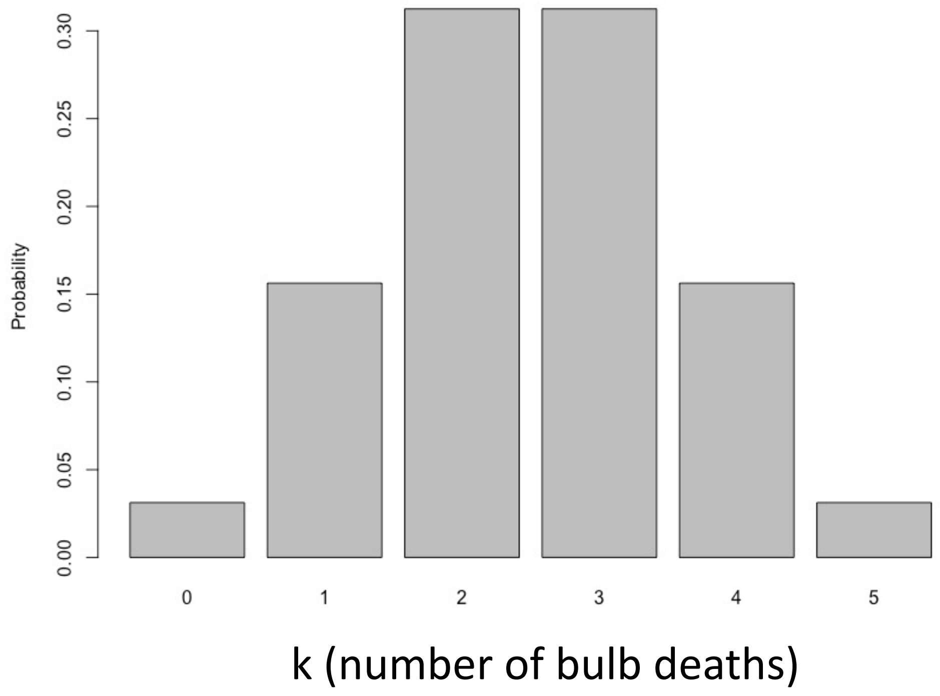
……………….…. all other possibilities

It's Binomial distribution

$$B(k = 2|n = 5, p = 0.5) = \binom{5}{2} 0.5^2 0.5^3$$

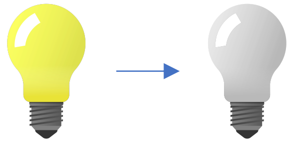
- Probability of *k* deaths is Binomial too

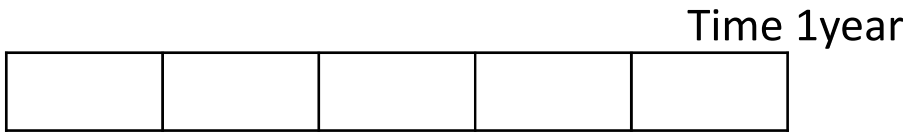
Number of bulb deaths over 1 year



Increasing number of bins

- 5 bins is a bad precision

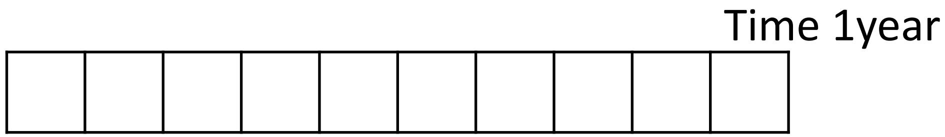




$\Delta t = 0.2\ year$

$$B_{old}(k|\, n = 5, p = 0.5) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Let's increase the number of bins to 10

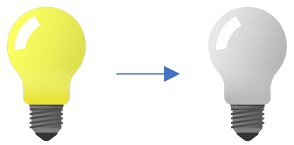


$\Delta t = 0.1\ year$

$$B_{new}(k|\, n = 10, p = 0.25) = \binom{n}{k} p^k (1-p)^{n-k}$$

Increasing number of bins

- 5 bins is a bad precision



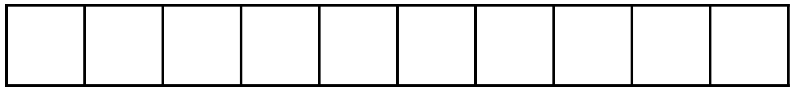
Time 1year

$\Delta t = 0.2\ year$

$$B_{old}(k|\, n = 5, p = 0.5) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Let's increase the number of bins to 10

Time 1year



$\Delta t = 0.1\ year$

$$B_{new}(k|\, n = 10, p = 0.25) = \binom{n}{k} p^k (1-p)^{n-k}$$

Key observations:

- Increasing number of bins ***n*** decreases ***p*** parameter
- Note, that the product ***np*** is the same in both cases
  - 10 * 0.25 = 5 * 0.5
- Let's denote this product as $\boldsymbol{\lambda = np}$

Motivation for Poisson distribution

- Binomial is not convenient for phenomena that continuously occur over time
- Let's re-write Binomial to make it "convenient"
- We use substitution: $\boldsymbol{\lambda = np}$
- **Mathematical trick:** note that $\boldsymbol{p = \lambda/n}$
- Take limit of $\boldsymbol{n}$ to get rid of the subjective split of time into bins

$$Binomial(k|n, \lambda/n) = \binom{n}{k} (\boldsymbol{\lambda/n})^k (1 - (\boldsymbol{\lambda/n}))^{n-k}$$

Motivation for Poisson distribution

- Binomial is not convenient for phenomena that continuously occur over time
- Let's re-write Binomial to make it "convenient"
- We use substitution: $\boldsymbol{\lambda = np}$
- **Mathematical trick:**
  - Note that $\boldsymbol{p = \lambda/n}$
  - Take limit of ***n*** to get rid of the subjective split of time into bins

$$Binomial(k|n, \lambda/n) = \binom{n}{k} (\lambda/n)^k (1 - (\lambda/n))^{n-k}$$

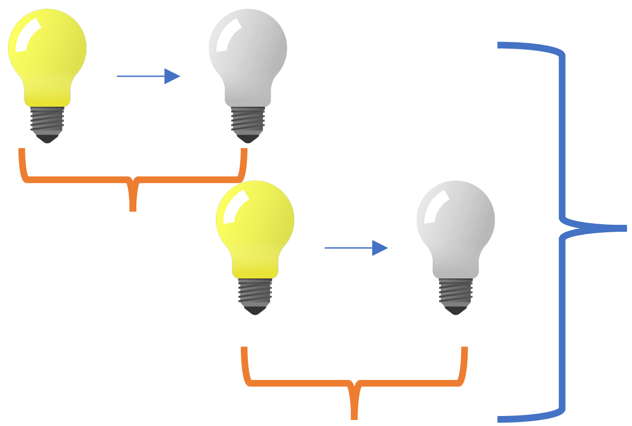
$$\text{Poisson}(k\ |\lambda) = \lim_{n->\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\boldsymbol{e^{-\lambda}\lambda^k}}{\boldsymbol{k}!}$$

$$\text{Poisson}(k\ |\lambda, t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

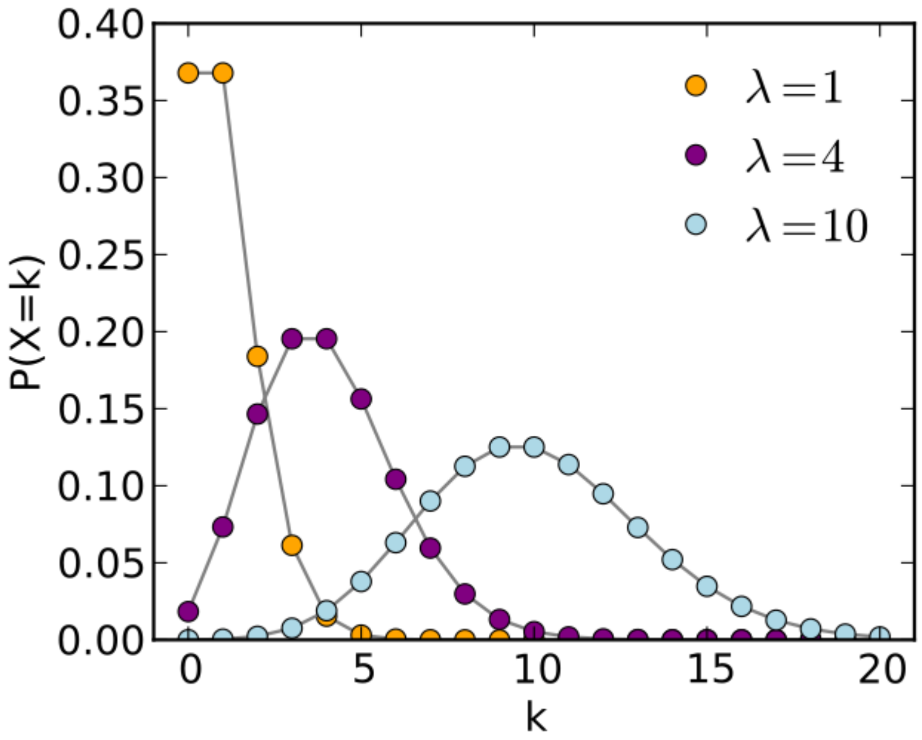
Poisson distribution

$$\text{Poisson}(k\,|\lambda, t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

- λ is called the rate parameter
- Poisson distr. shows the number of changes k given λ and time t



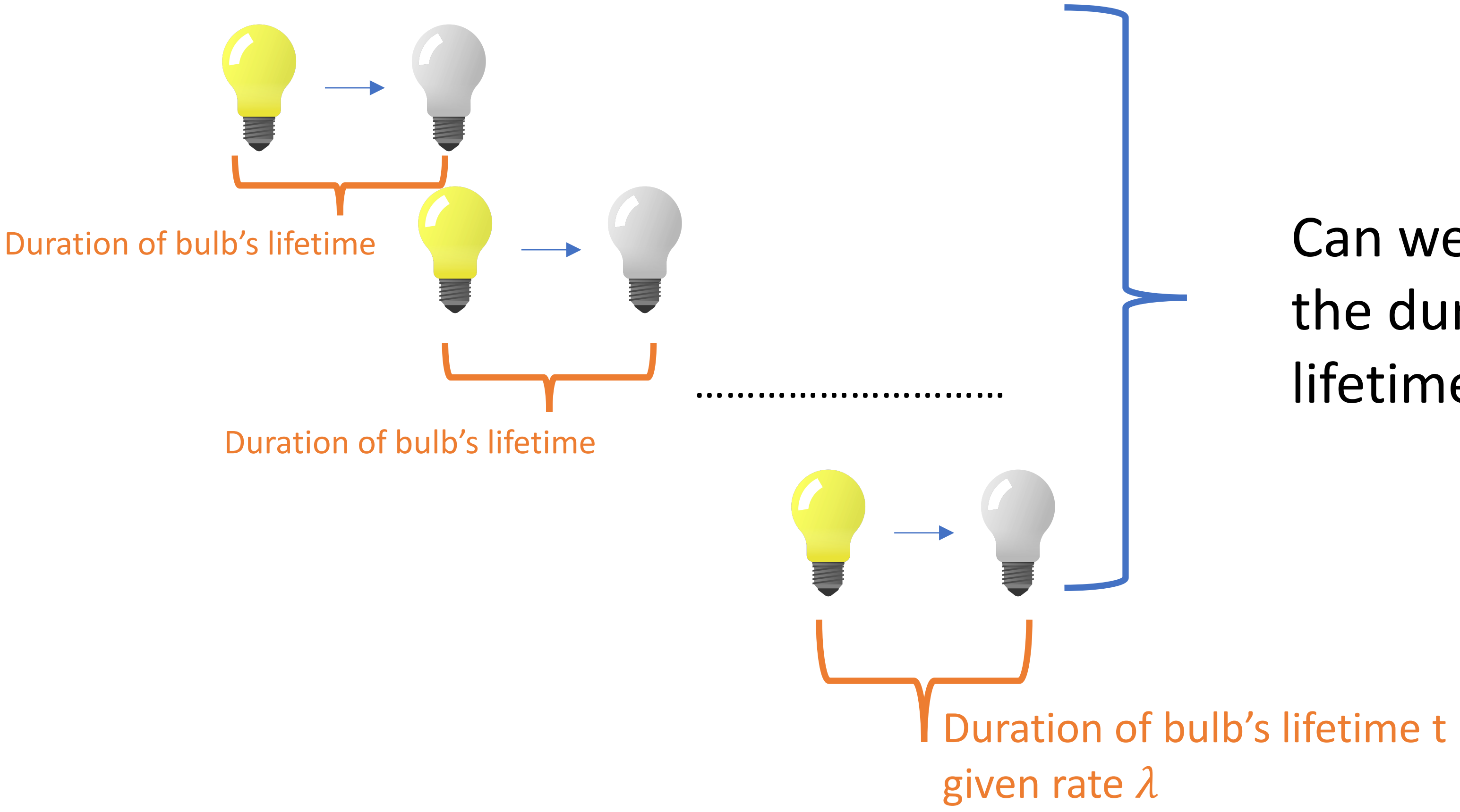
Number of dead bulb's per year



Ingredients to derive continuous-time Markov models

Binomial distribution → Poisson distrubution → Exponential distribution → Continuous-time and discrete-state Markov models

From Poisson to Exponential distribution



Can we derive a distribution for the duration of the bulb’s lifetime?

Let’s step aside and think how this distribution would look like

Deriving Exponential distribution

$$\text{Poisson}(k \mid \lambda, t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Observing exactly 0 changes

$$\text{Poisson}(k = 0 \mid \lambda, t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

Observing more than 0 changes (something happens)

$$\text{Poisson}(k > 0 \mid \lambda, t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = 1 - e^{-\lambda t}$$

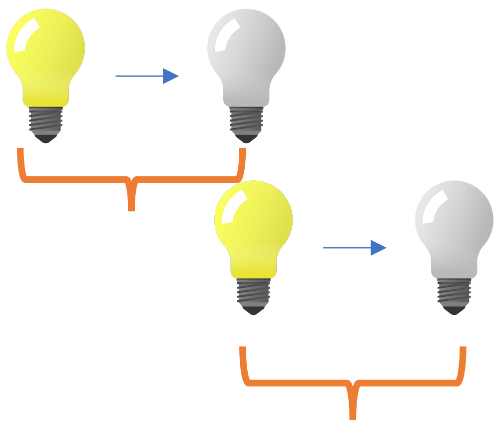
Differentiating with respect to time

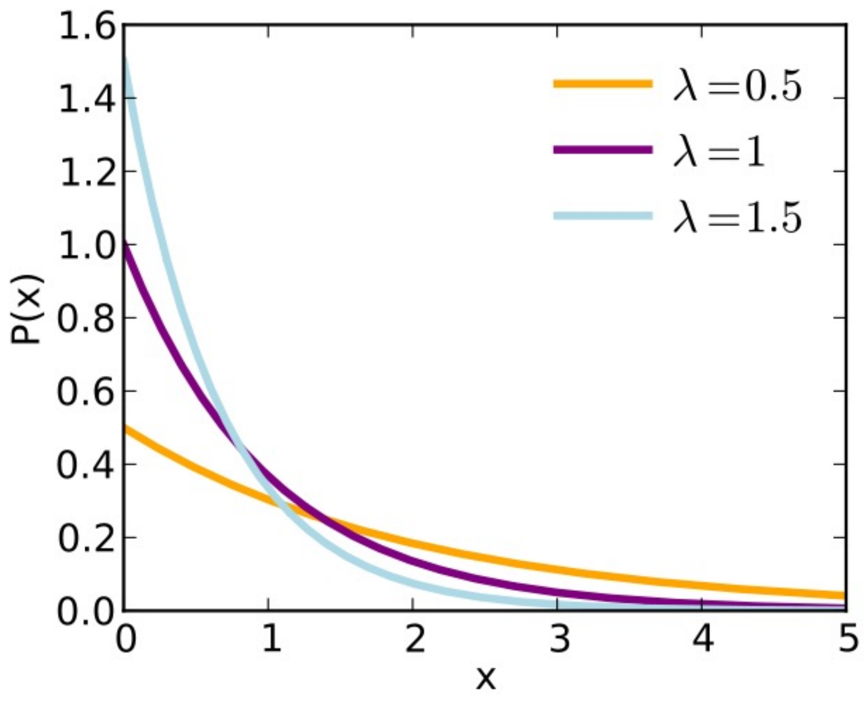
$$\frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

$$\text{Exponential}(\, t \mid \lambda) = \lambda e^{-\lambda t}$$

Exponential distribution

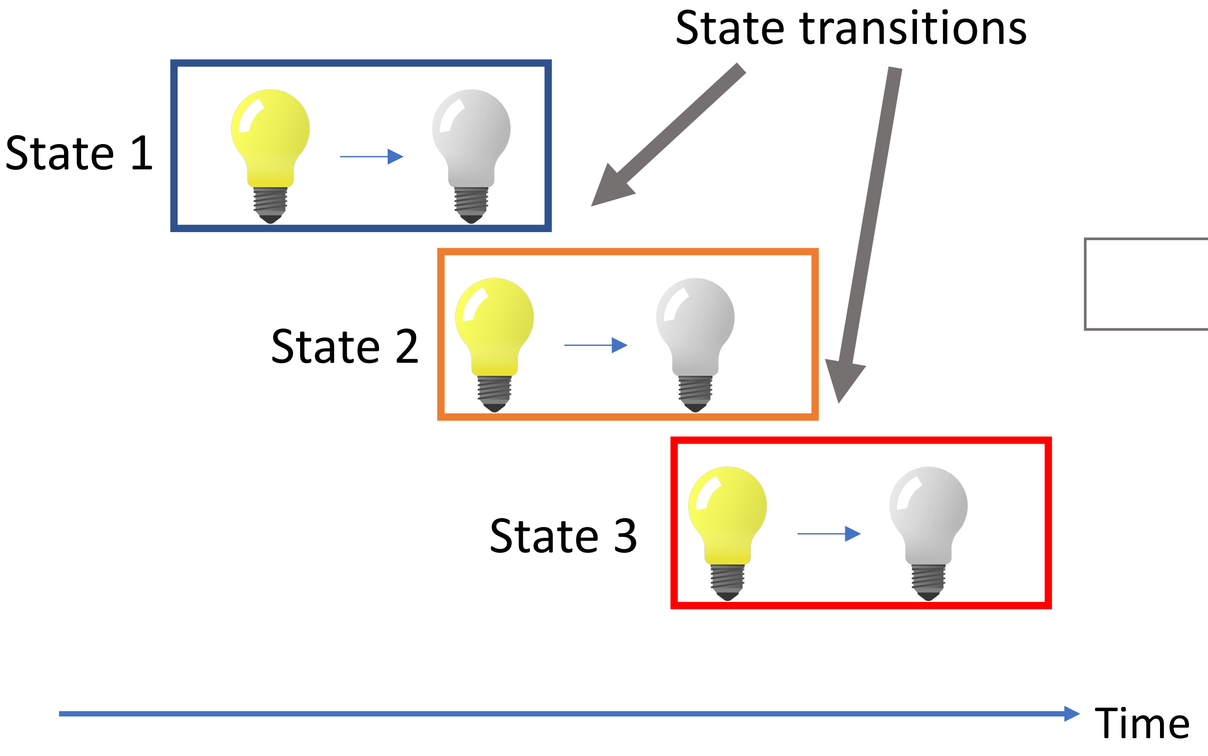
$$\text{Exponential}\,(\,t\mid\lambda) = \lambda e^{-\lambda t}$$



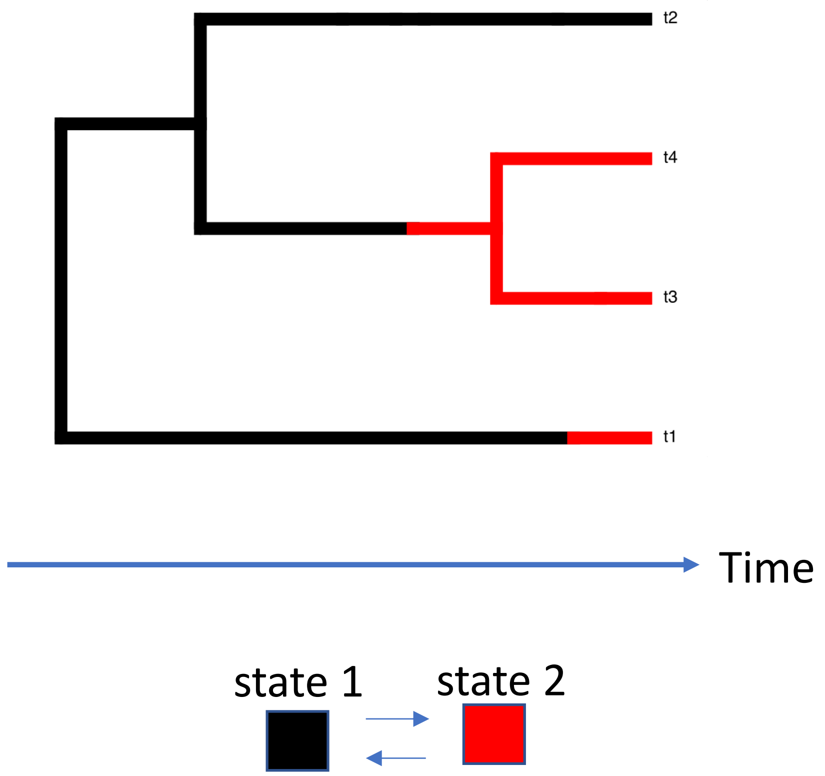


- Exponential and Poisson are the same processes but different aspects
- Same interpretation of the parameter λ **(=rate)**
- λ is the mean number of changes over time interval in Poisson

Now we can model events occurring over time !



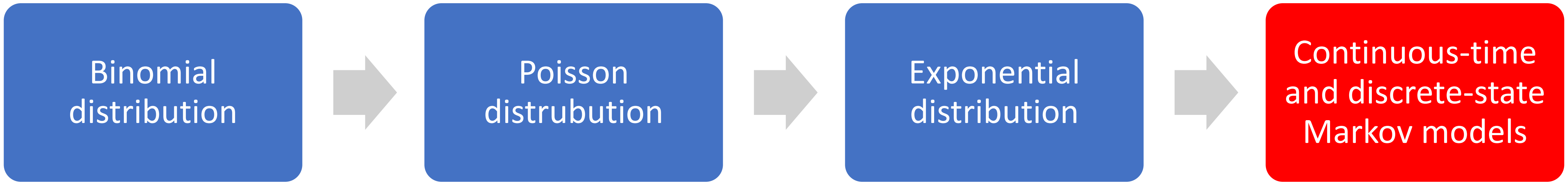
Evolution of characters on a tree is the the state transitions over time



$$\text{Poisson}(k \mid \lambda, t) = \frac{e^{-\lambda t} \lambda t^{k}}{k!}$$

$$\text{Exponential}(t \mid \lambda) = \lambda e^{-\lambda t}$$

Ingredients to derive continuous-time Markov models

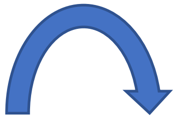


Modeling phylogenetic process: Markov models (Markov chains)

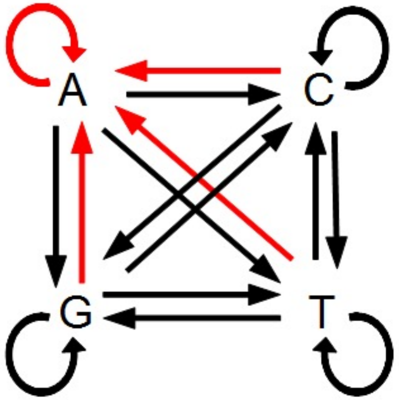
- In probability theory, a Markov model is a stochastic model used to model randomly changing systems
- It is assumed that future states depend only on the current state, not on the events that occurred before it (that is, it assumes the Markov property)

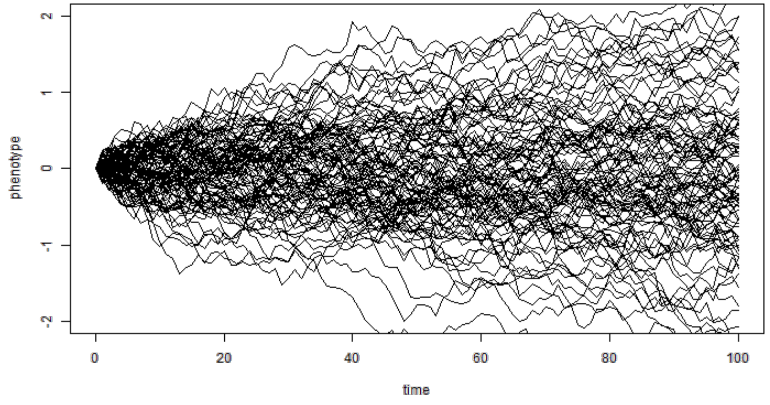
Andrey Markov (1856 – 1922)

Future state depends only on present state



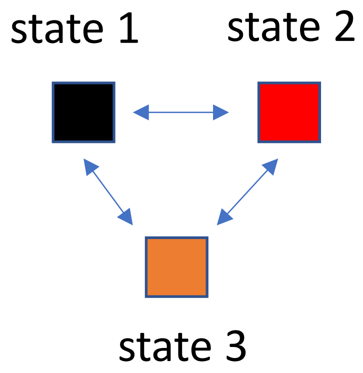
past state –> present state –> future state



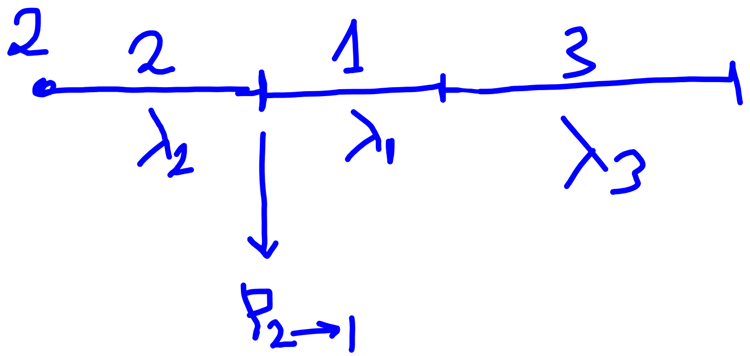


Continuous-time Markov models: creating transition rate matrix

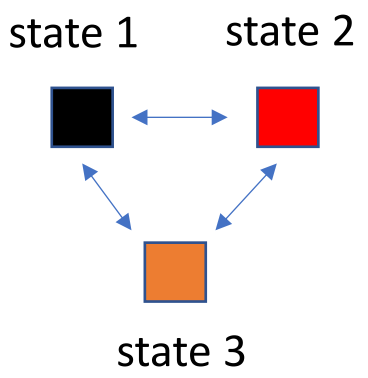
- Let’s generalize exponential distribution for modeling transitions between discrete states
- Let’s assume that we have a system (organism) that come in three states
- We that the waiting time of staying in each state is exponential distribution



Representing Markov chain evolution



Continuous-time Markov models: creating transition rate matrix



- Let’s represent our evolving system in a smart way
- Main diagonal elements are rates from the exponential distribution

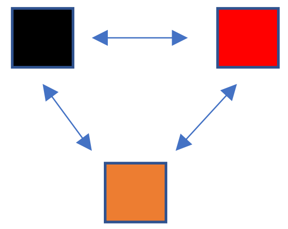
To:

From:

$$\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & & \\ & 1 & \\ & & 1.2 \end{bmatrix}$$

Continuous-time Markov models: creating transition rate matrix

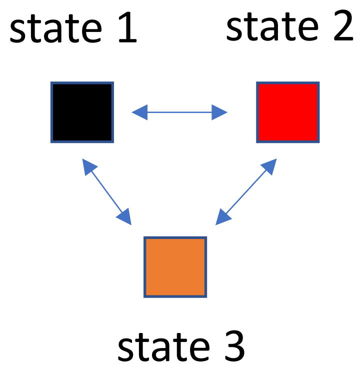


- Let's represent our evolving system in a smart way
- Main diagonal elements are rates from the exponential distribution
- Off-diagonal elements are probabilities

$$\begin{bmatrix} \lambda_1 & p_{12} & p_{13} \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & 0.8 & 0.2 \\ & 1 & \\ & & 1.2 \end{bmatrix}$$

Continuous-time Markov models: creating transition rate matrix

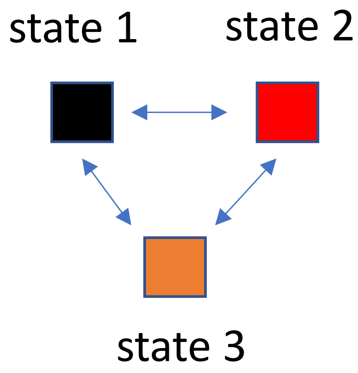


- Let’s represent our evolving system in a smart way
- Main diagonal elements are rates from the exponential distribution
- Off-diagonal elements are probabilities

$$\begin{bmatrix} \lambda_1 & p_{12} & p_{13} \\ p_{21} & \lambda_2 & p_{23} \\ p_{31} & p_{32} & \lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & 0.8 & 0.2 \\ 0.8 & 1 & 0.2 \\ 0.8 & 0.2 & 1.2 \end{bmatrix}$$

Continuous-time Markov models: creating transition rate matrix

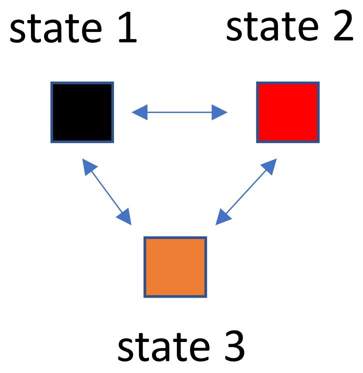


- For mathematical convenience let's rescale probabilities by rates

$$\begin{bmatrix} \lambda_1 & \lambda_1 * p_{12} & \lambda_1 * p_{13} \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & 0.5 * 0.8 & 0.5 * 0.2 \\ & 1 & \\ & & 1.2 \end{bmatrix}$$

Continuous-time Markov models: creating transition rate matrix

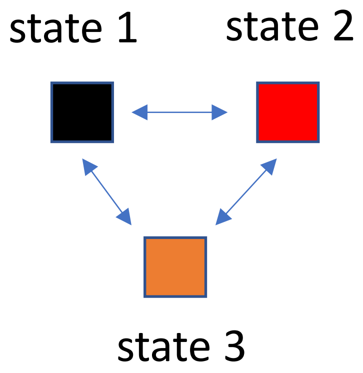


- For mathematical convenience let's make the rates negative

$$\begin{bmatrix} -\lambda_1 & \lambda_1 * p_{12} & \lambda_1 * p_{13} \\ & -\lambda_2 & \\ & & -\lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} -0.5 & 0.5 * 0.8 & 0.5 * 0.2 \\ & -1 & \\ & & -1.2 \end{bmatrix}$$

Continuous-time Markov models: creating transition rate matrix



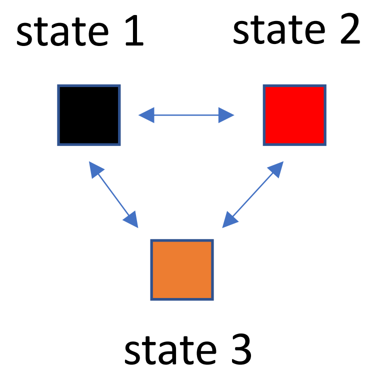
$$\begin{bmatrix} \lambda_1 & p_1 & p_2 \\ p_1 & \lambda_2 & p_2 \\ p_1 & p_2 & \lambda_3 \end{bmatrix} \quad \begin{bmatrix} 0.5 & 0.8 & 0.2 \\ 0.8 & 1 & 0.2 \\ 0.8 & 0.2 & 1.2 \end{bmatrix}$$

Non-rescaled matrix

$$\begin{bmatrix} -\lambda_1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & -\lambda_2 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & -\lambda_3 \end{bmatrix} \quad \begin{bmatrix} -0.5 & 0.4 & 0.1 \\ 0.8 & -1 & 0.2 \\ 0.96 & 0.24 & -1.2 \end{bmatrix}$$

Transition rate matrix.
Rescaled matrix with entities called infinitesimal rates

From rates to probabilities



- **Transition rate matrix.** Infinitesimal rates

$$\boldsymbol{Q} = \begin{bmatrix} -0.5 & 0.4 & 0.1 \\ 0.8 & -1 & 0.2 \\ 0.96 & 0.24 & -1.2 \end{bmatrix}$$

- **Probability transition matrix.** Exponentiate rate matrix

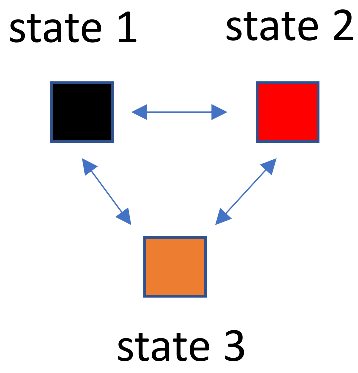
$$P(\boldsymbol{Q}, t) = e^{Qt}$$

$$e^{Q*1} = \begin{bmatrix} 0.72 & 0.2 & 0.08 \\ 0.46 & 0.46 & 0.08 \\ 0.46 & 0.2 & 0.34 \end{bmatrix}$$

Matrix exponential transforms rates into probabilities:

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

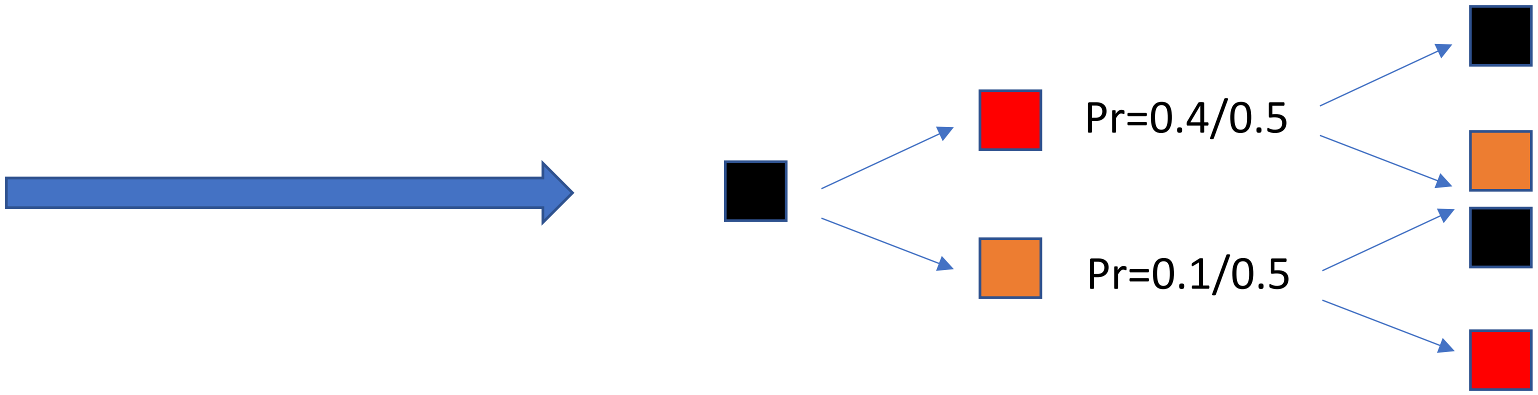
Continuous-time Markov models: creating transition rate matrix



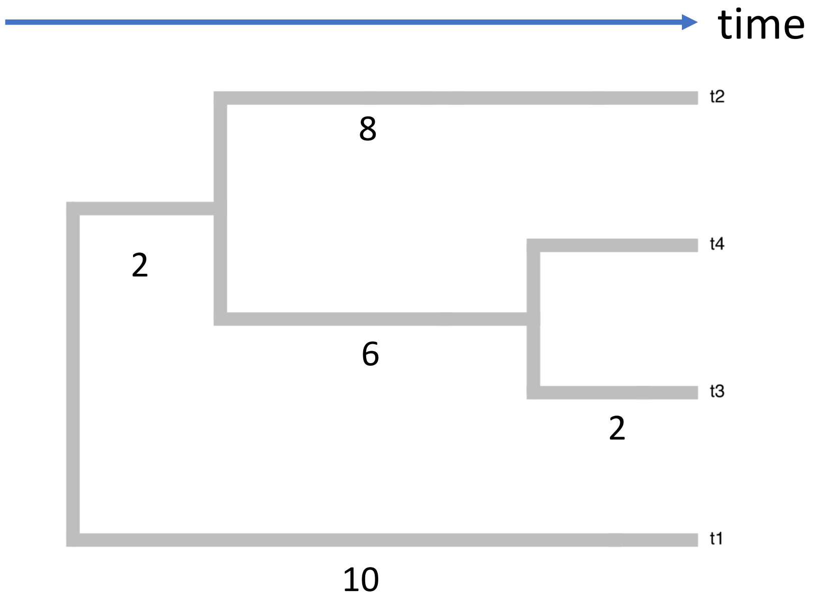
$$e^{Q*1} = \begin{bmatrix} 0.72 & 0.2 & 0.08 \\ 0.46 & 0.46 & 0.08 \\ 0.46 & 0.2 & 0.34 \end{bmatrix} \qquad P(\boldsymbol{Q}, t) = e^{Qt}$$

Initial vector of probabilities

(1/3, 1/3, 1/3)

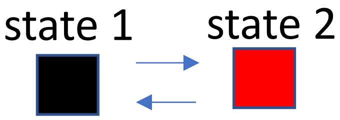


Simulating data under Markov models on a tree

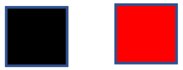


Random number generator

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

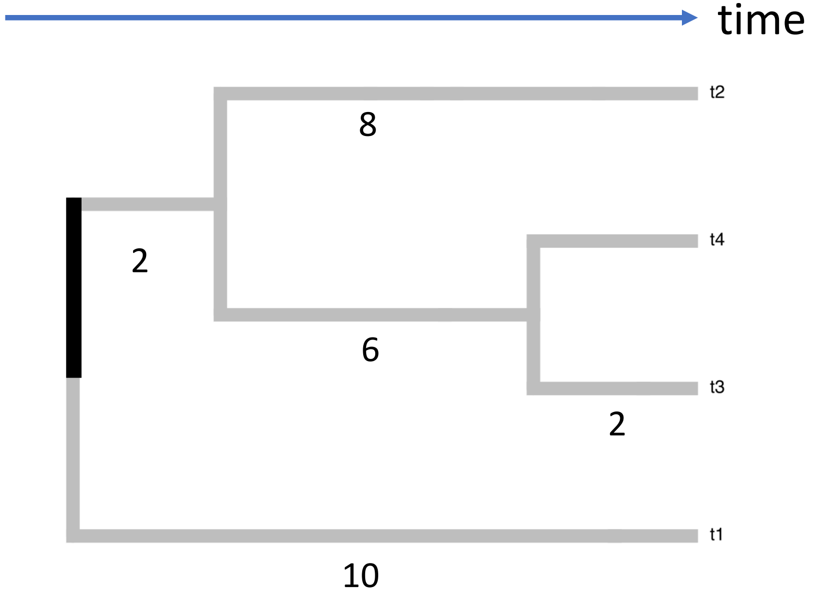


Initial vector



$\pi = (1/2, 1/2)$

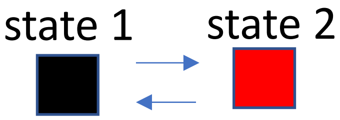
Simulating data under Markov models on a tree



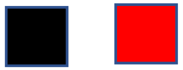
Random number generator

1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

$$\boldsymbol{Q}=\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

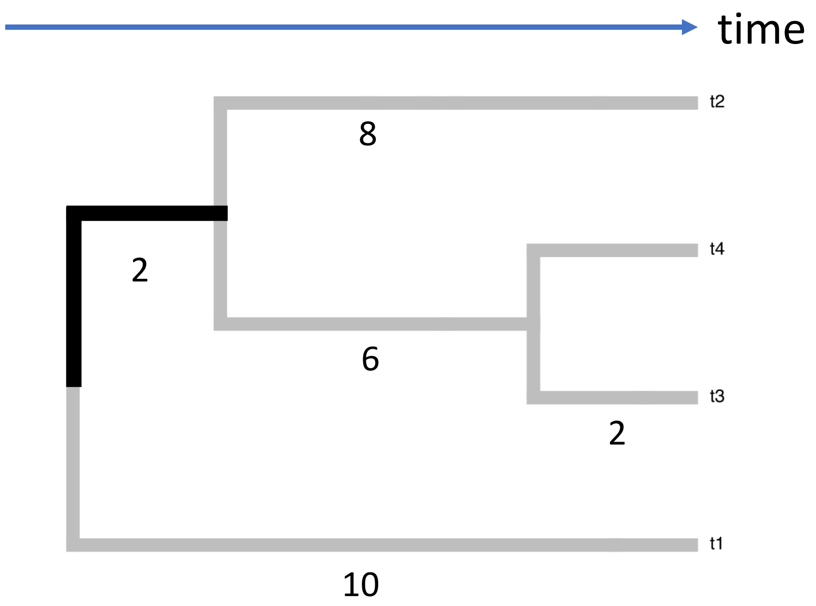


Initial vector



π =(1/2, 1/2)

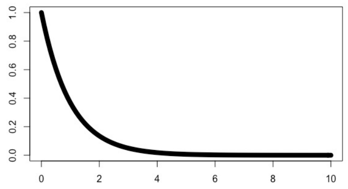
Simulating data under Markov models on a tree



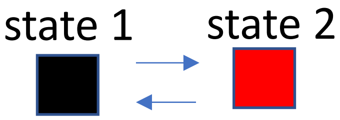
Random number generator

1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4

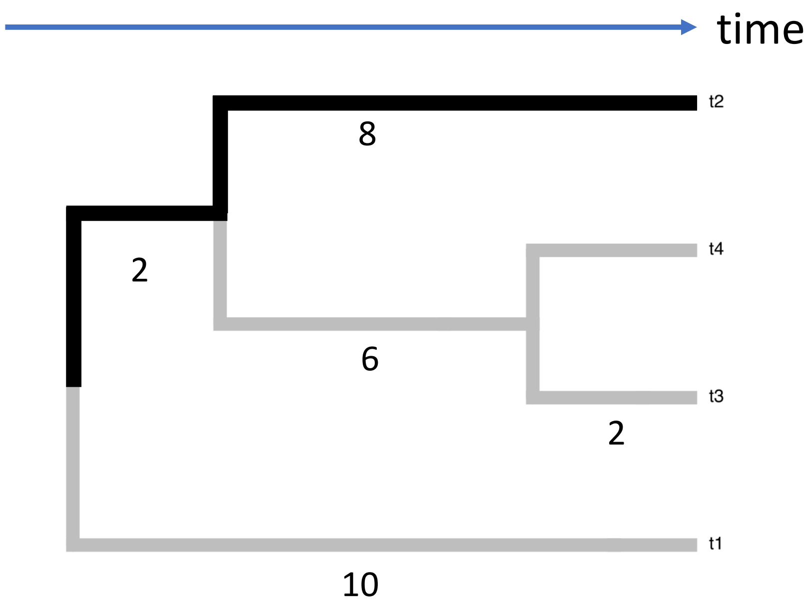


$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$



Initial vector $\pi = (1/2, 1/2)$

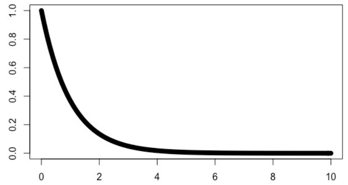
Simulating data under Markov models on a tree



Random number generator

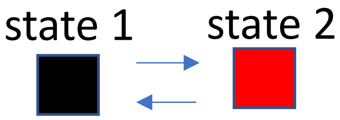
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4

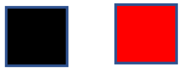


3. Draw a random number from Exp($\lambda = 1$). RND=8.4

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

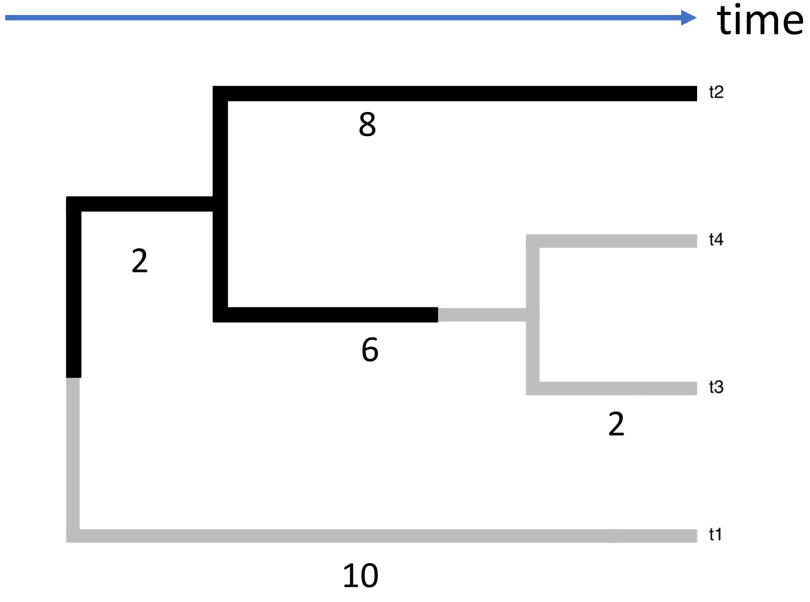


Initial vector



π =(1/2, 1/2)

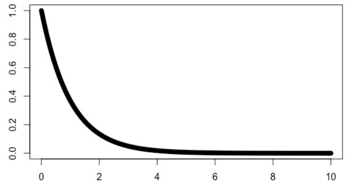
Simulating data under Markov models on a tree



Random number generator

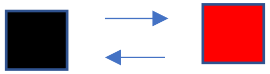
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4

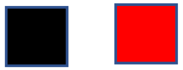


3. Draw a random number from Exp($\lambda = 1$). RND=8.4
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)

$\boldsymbol{Q}=\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$ state 1 state 2

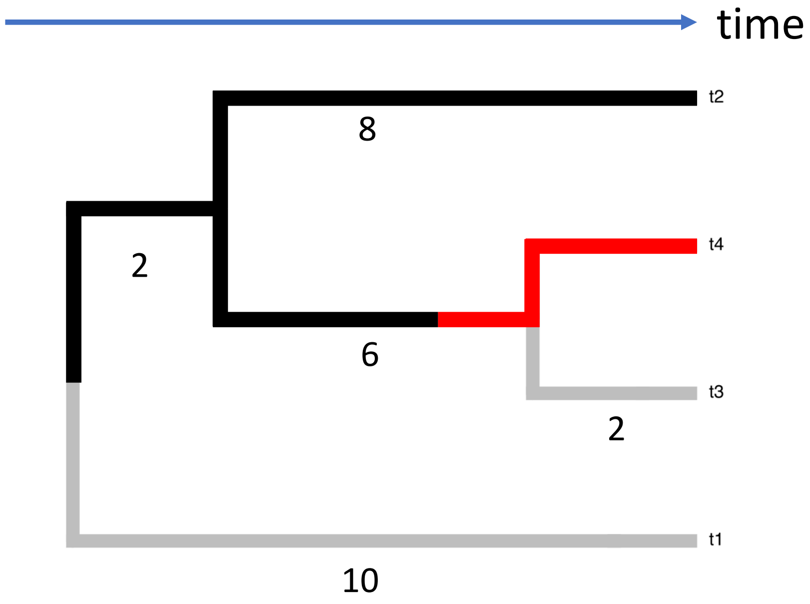


Initial vector



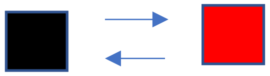
π =(1/2, 1/2)

Simulating data under Markov models on a tree

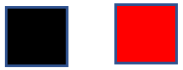


$$\mathbf{Q}=\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

state 1 state 2



Initial vector

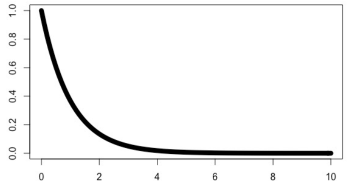


π =(1/2, 1/2)

Random number generator

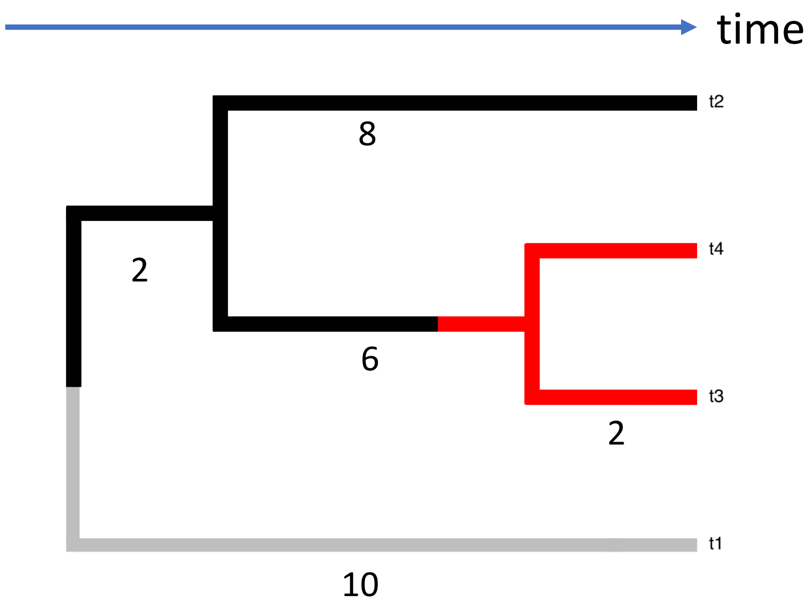
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4



3. Draw a random number from Exp($\lambda = 1$). RND=8.4
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)
5. Draw a random number from Exp($\lambda = 2$). RND=4.6

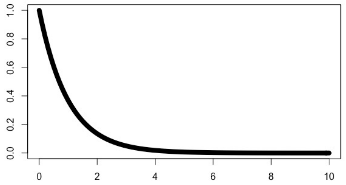
Simulating data under Markov models on a tree



Random number generator

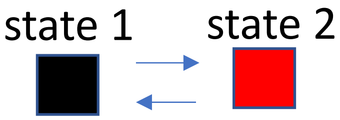
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4

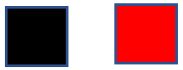


3. Draw a random number from Exp($\lambda = 1$). RND=8
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)
5. Draw a random number from Exp($\lambda = 2$). RND=4.6
6. Draw a random number from Exp($\lambda = 2$). RND=4.9

$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$

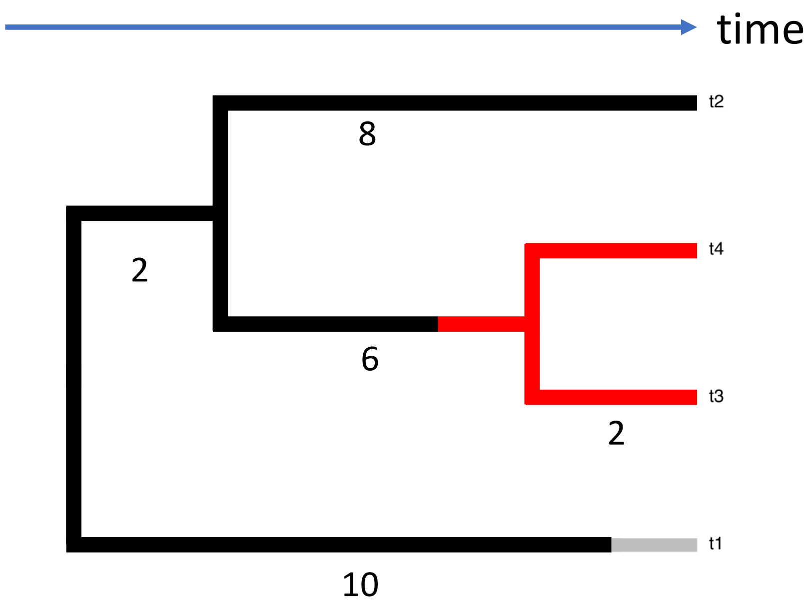


Initial vector

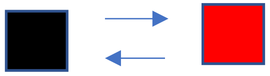


$\pi = (1/2, 1/2)$

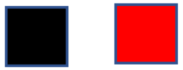
Simulating data under Markov models on a tree



$$Q=\begin{bmatrix}-1 & 1\\ 2 & -2\end{bmatrix}$$



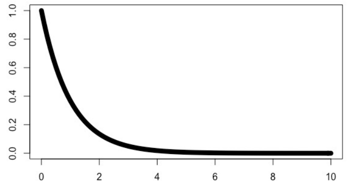
Initial vector



π =(1/2, 1/2)

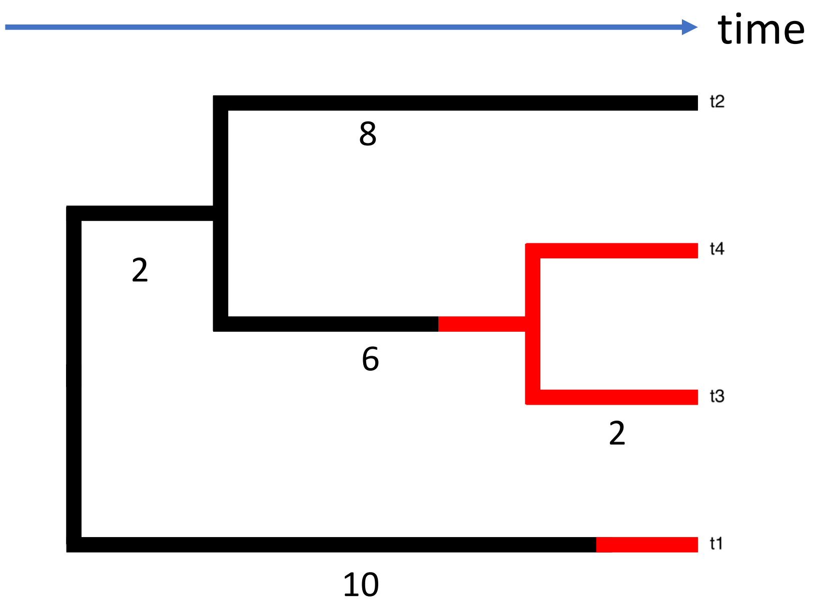
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4



3. Draw a random number from Exp($\lambda = 1$). RND=8
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)
5. Draw a random number from Exp($\lambda = 2$). RND=4.6
6. Draw a random number from Exp($\lambda = 2$). RND=4.9
7. Draw a random number from Exp($\lambda = 1$). RND=9.1 (to state 2)

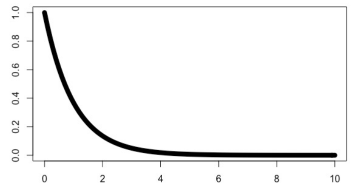
Simulating data under Markov models on a tree



Random number generator

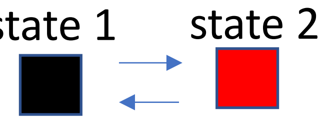
1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4



3. Draw a random number from Exp($\lambda = 1$). RND=8
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)
5. Draw a random number from Exp($\lambda = 2$). RND=4.6
6. Draw a random number from Exp($\lambda = 2$). RND=4.9
7. Draw a random number from Exp($\lambda = 1$). RND=9.1 (to state 2)
8. Draw a random number from Exp($\lambda = 2$). RND=3.3

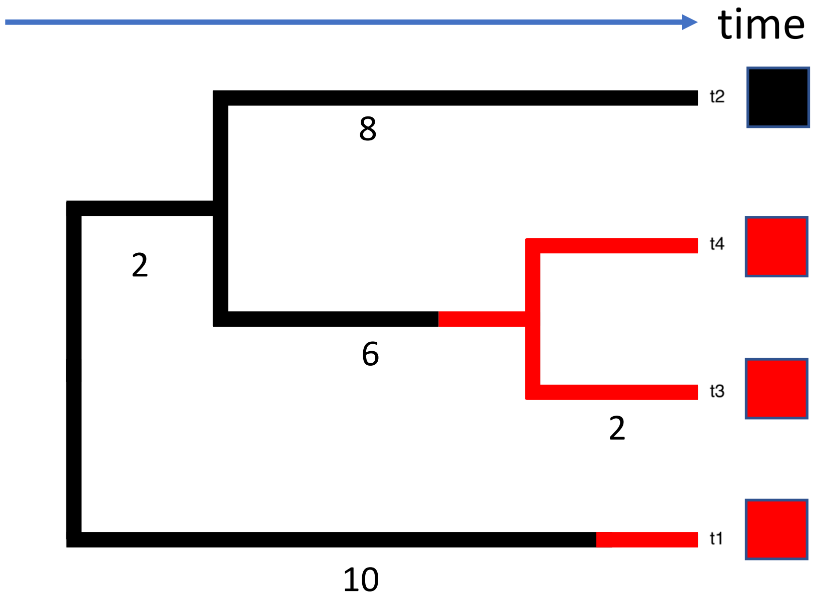
$$\boldsymbol{Q}=\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$



Initial vector

π =(1/2, 1/2)

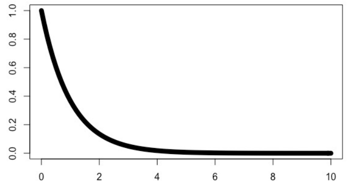
Simulating data under Markov models on a tree



Random number generator

1. Randomly select state at the root from a uniform distribution. RND=0.4 (starting state 1)

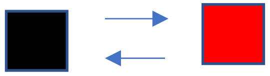
2. Draw a random number from Exponential distribution with $\lambda = 1$. RND=2.4



3. Draw a random number from Exp($\lambda = 1$). RND=8
4. Draw a random number from Exp($\lambda = 1$). RND=4.2 (to state 2)
5. Draw a random number from Exp($\lambda = 2$). RND=4.6
6. Draw a random number from Exp($\lambda = 2$). RND=4.9
7. Draw a random number from Exp($\lambda = 1$). RND=9.1 (to state 2)
8. Draw a random number from Exp($\lambda = 2$). RND=3.3

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

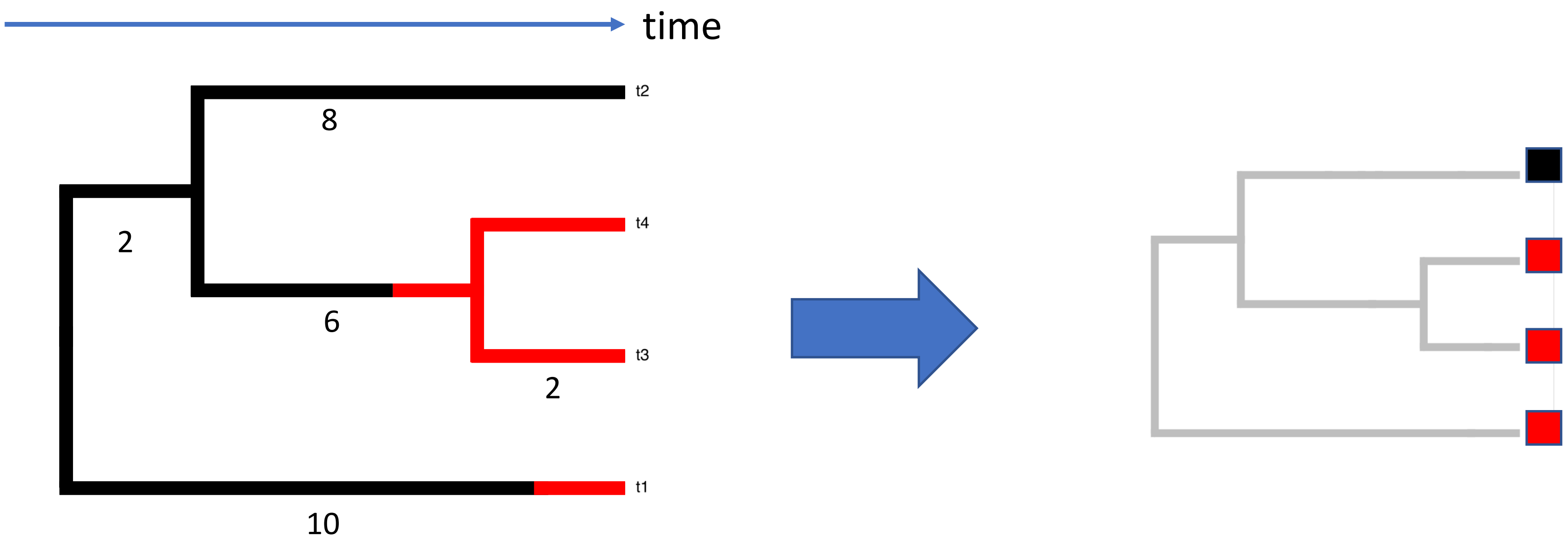
state 1 state 2



Initial vector

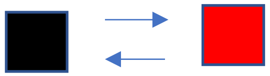
π =(1/2, 1/2)

Now let’s run likelihood inference

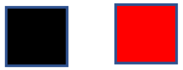


$$\boldsymbol{Q}=\begin{bmatrix}-1 & 1\\ 2 & -2\end{bmatrix}$$

state 1 state 2

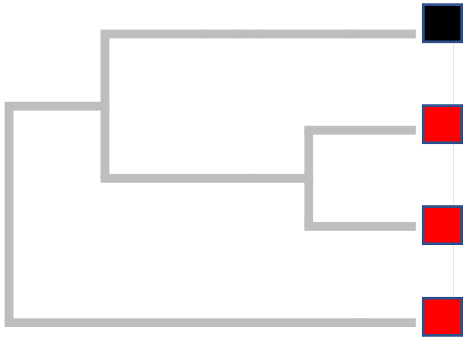


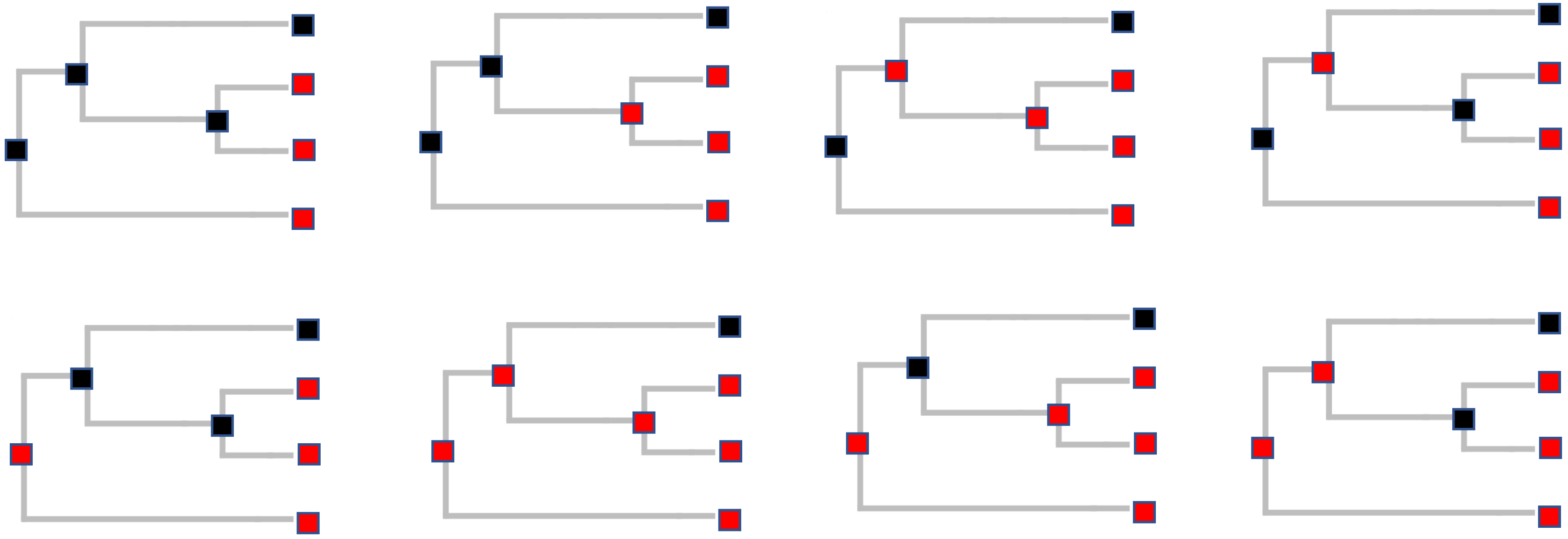
Initial vector



π =(1/2, 1/2)

Inference: estimating tree likelihood

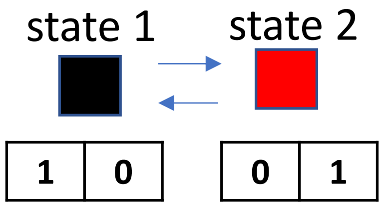




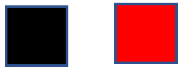
Integrate over all possible combinations of the ancestral states

Felsenstein's pruning algorithm

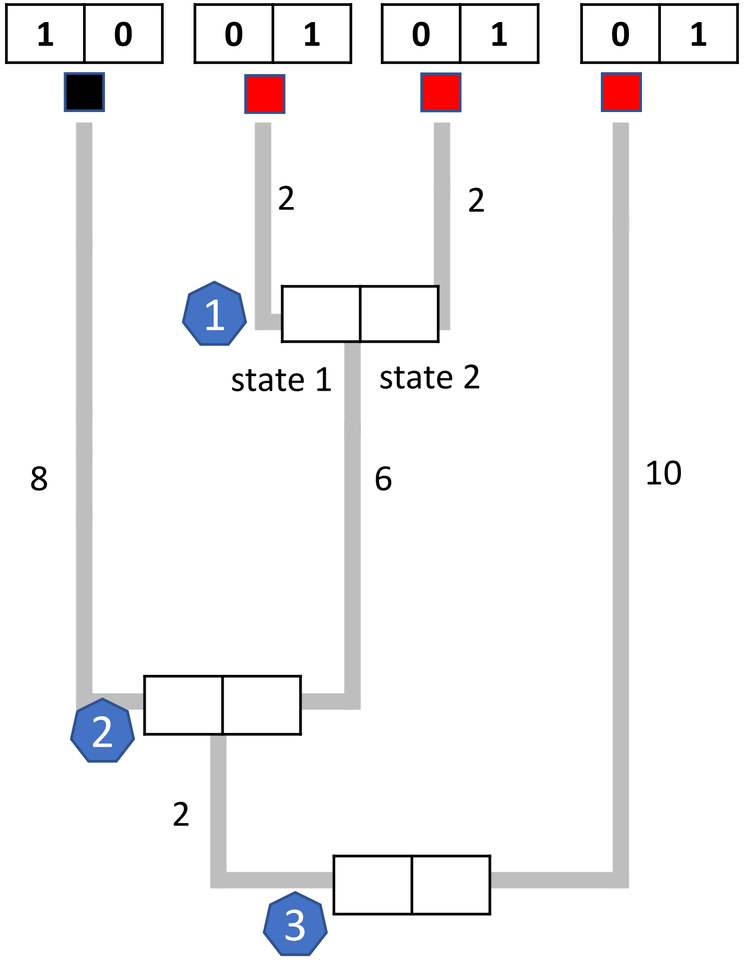
Given values:



$$\boldsymbol{Q}=\begin{bmatrix}-1 & 1\\ 2 & -2\end{bmatrix}$$



π =(1/2, 1/2)



Summary

- We have derived a discrete state Markov model from the chain consisting of Binomial, Poisson and Exponential distributions.
- Discrete state Markov model is the core of almost all phylogenetic approaches that use different type of data (morphology, DNA, proteins, etc.)