

---

# Introduction to Phylogenetics

## Part 3

Dirk Husmeier

Biomathematics and Statistics Scotland  
at the Scottish Crop Research Institute  
Invergowrie, Dundee DD2 5DA, UK

Email: [dirk@bioss.ac.uk](mailto:dirk@bioss.ac.uk)

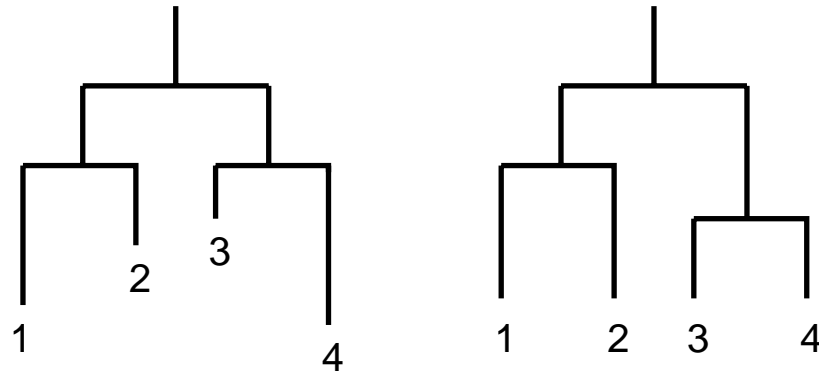
<http://www.bioss.ac.uk/~dirk>

---

## Testing the molecular clock hypothesis

---

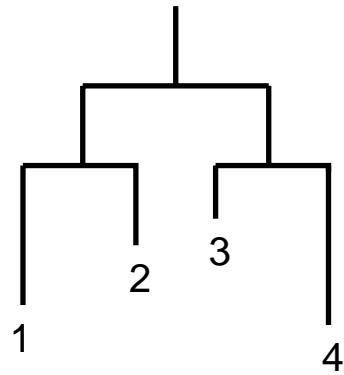
$H_0$  : ultrametric tree (molecular clock)  
( $K - 1$  constraints,  $K =$  number of leaf nodes).



## Testing the molecular clock hypothesis

---

$H_0$  : ultrametric tree (molecular clock)  
( $K - 1$  constraints,  $K$  = number of leaf nodes).



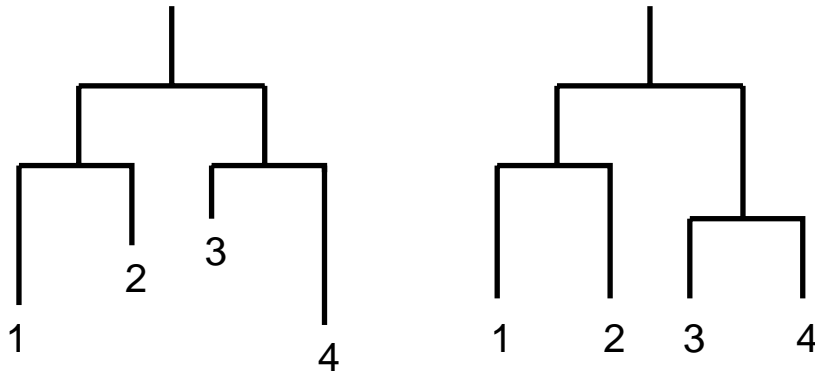
$$L_1 = \ln P(D|\mathbf{w}_1, H_1)$$

$$L_0 = \ln P(D|\mathbf{w}_0, H_0)$$

## Testing the molecular clock hypothesis

---

$H_0$  : ultrametric tree (molecular clock)  
( $K - 1$  constraints,  $K =$  number of leaf nodes).



$$L_1 = \ln P(D|\mathbf{w}_1, H_1) \quad L_0 = \ln P(D|\mathbf{w}_0, H_0)$$

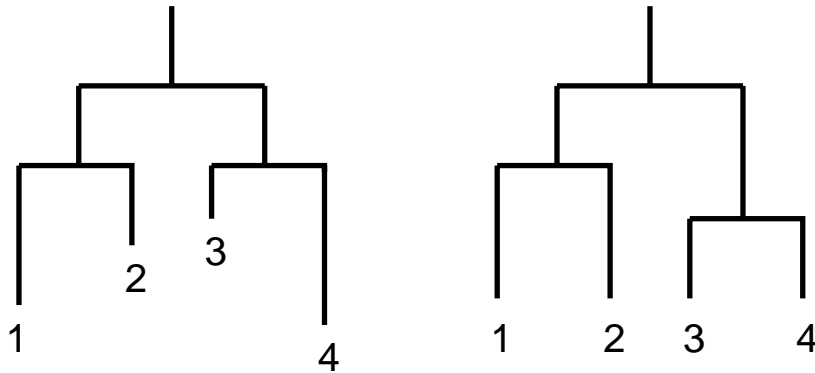
$L_1 > L_0$ , but is this **significant?**

---

## Testing the molecular clock hypothesis

---

$H_0$  : ultrametric tree (molecular clock)  
( $K - 1$  constraints,  $K$  = number of leaf nodes).



$$L_1 = \ln P(D|\mathbf{w}_1, H_1) \quad L_0 = \ln P(D|\mathbf{w}_0, H_0)$$

$L_1 > L_0$ , but is this **significant?**

Nested models

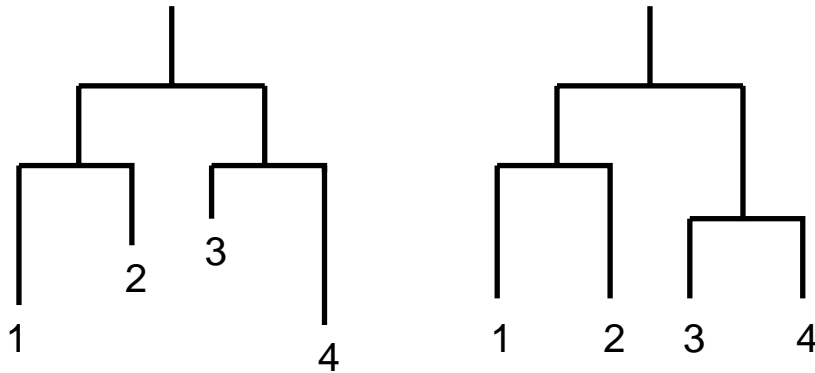
Likelihood ratio test:  $2(L_1 - L_0) \sim \chi^2_{(K-1)}$

---

## Testing the molecular clock hypothesis

---

$H_0$  : ultrametric tree (molecular clock)  
( $K - 1$  constraints,  $K =$  number of leaf nodes).



$$L_1 = \ln P(D|\mathbf{w}_1, H_1) \quad L_0 = \ln P(D|\mathbf{w}_0, H_0)$$

$L_1 > L_0$ , but is this **significant?**

Nested models

**Likelihood ratio test:**  $2(L_1 - L_0) \sim \chi^2_{(K-1)}$

But only valid for  $N \rightarrow \infty$

---

## Parametric bootstrapping

---

- Generate the **null distribution** using **computer simulations** .
- Overcome the restriction to **asymptotic results** .

## Parametric bootstrapping

---

- Generate the **null distribution** using **computer simulations** .
- Overcome the restriction to **asymptotic results** .
- Estimate both models from the data with ML, compute  
 $\Delta L = L_1 - L_0$ .

## Parametric bootstrapping

---

- Generate the **null distribution** using **computer simulations** .
- Overcome the restriction to **asymptotic results** .
- Estimate both models from the data with ML, compute  $\Delta L = L_1 - L_0$ .
- Simulate synthetic data  $D_1, D_2, \dots, D_B$  from the constrained model,  $H_0$ .

## Parametric bootstrapping

---

- Generate the **null distribution** using **computer simulations** .
  - Overcome the restriction to **asymptotic results** .
  - Estimate both models from the data with ML, compute  $\Delta L = L_1 - L_0$ .
  - Simulate synthetic data  $D_1, D_2, \dots, D_B$  from the constrained model,  $H_0$ .
  - For each synthetic data set, estimate both models with ML and compute  $\Delta L^i = L_1^i - L_0^i$
-

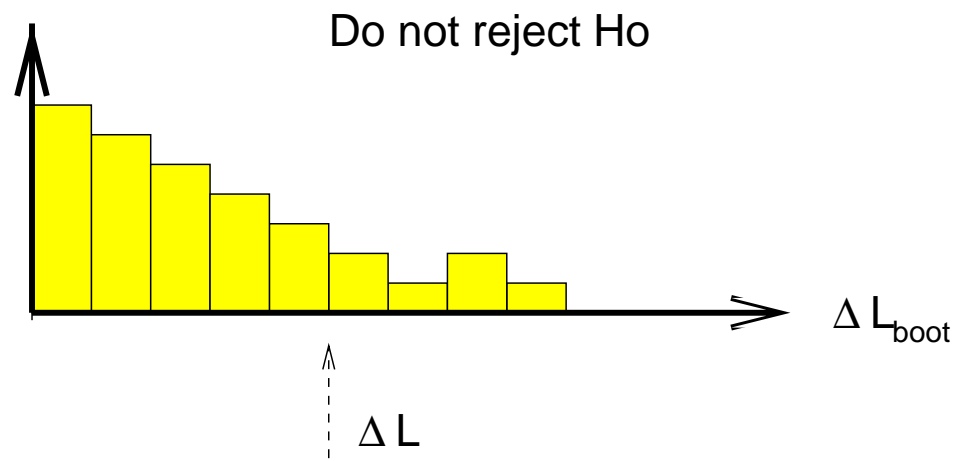
## Parametric bootstrapping

---

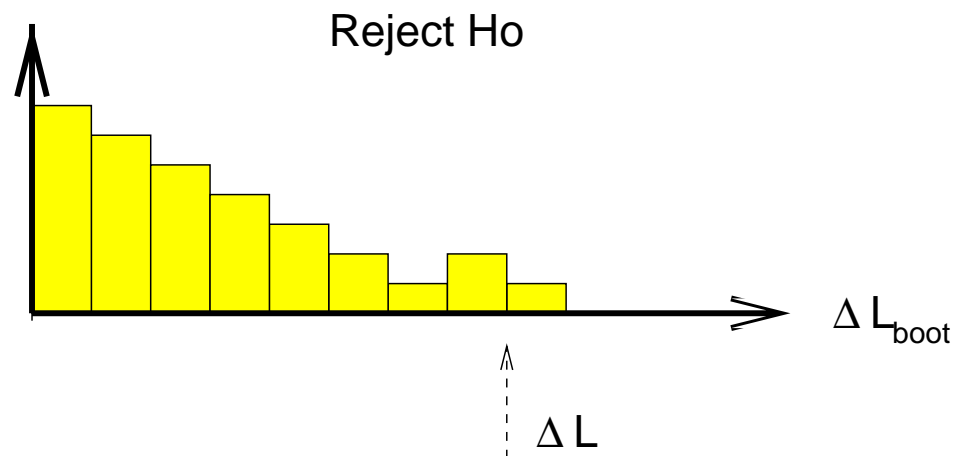
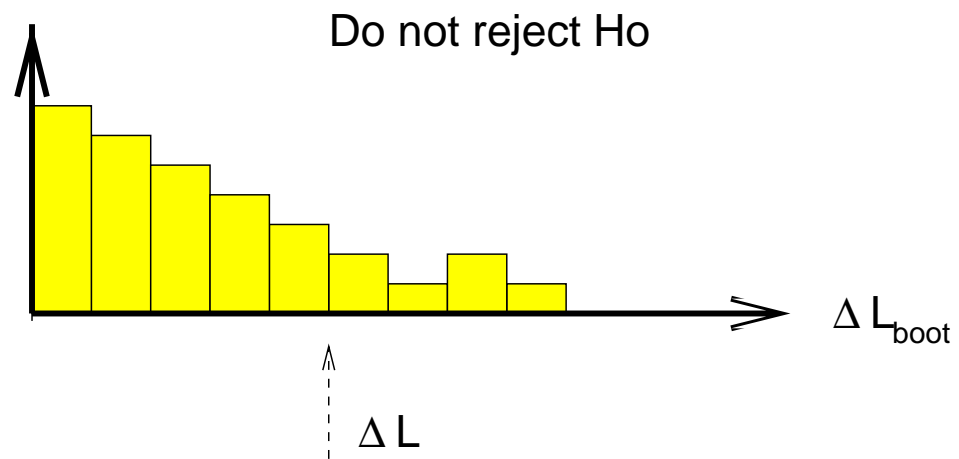
- Generate the **null distribution** using **computer simulations** .
  - Overcome the restriction to **asymptotic results** .
  - Estimate both models from the data with ML, compute  $\Delta L = L_1 - L_0$ .
  - Simulate synthetic data  $D_1, D_2, \dots, D_B$  from the constrained model,  $H_0$ .
  - For each synthetic data set, estimate both models with ML and compute  $\Delta L^i = L_1^i - L_0^i$
  - From this **null distribution** , get the **P-value** :  
 $|\{i | \Delta L^i > \Delta L\}| = m, \Rightarrow \text{P-value} = m/B.$
-

# Parametric bootstrapping

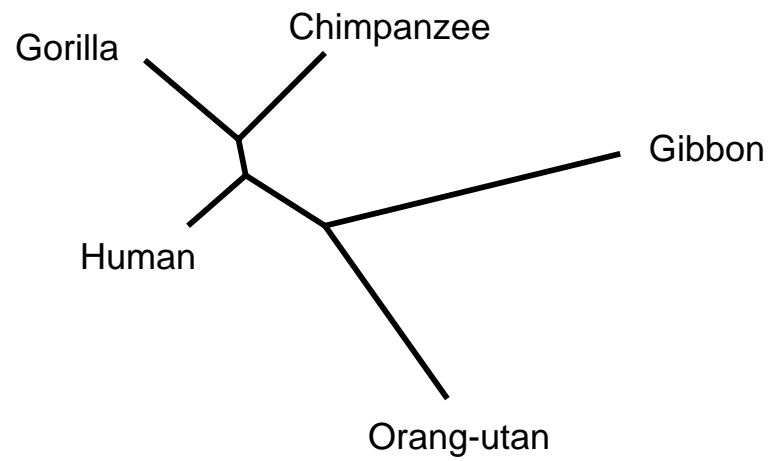
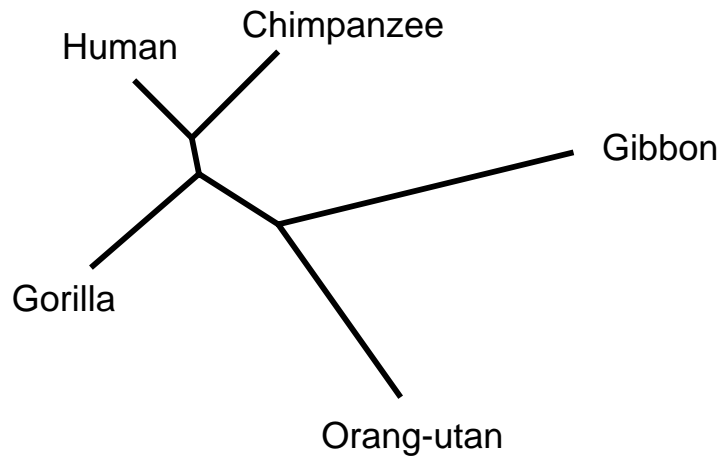
---



## Parametric bootstrapping

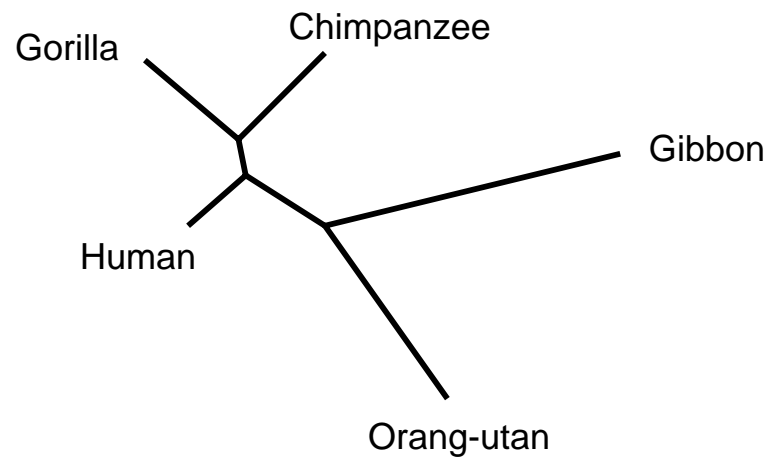
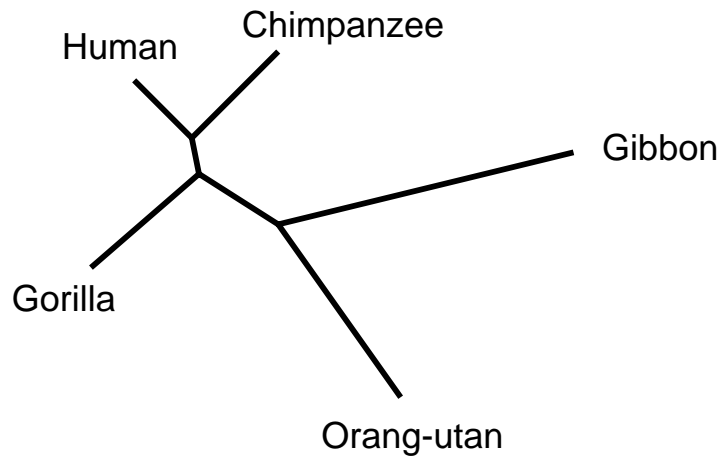


## Non-nested models



## Non-nested models

---



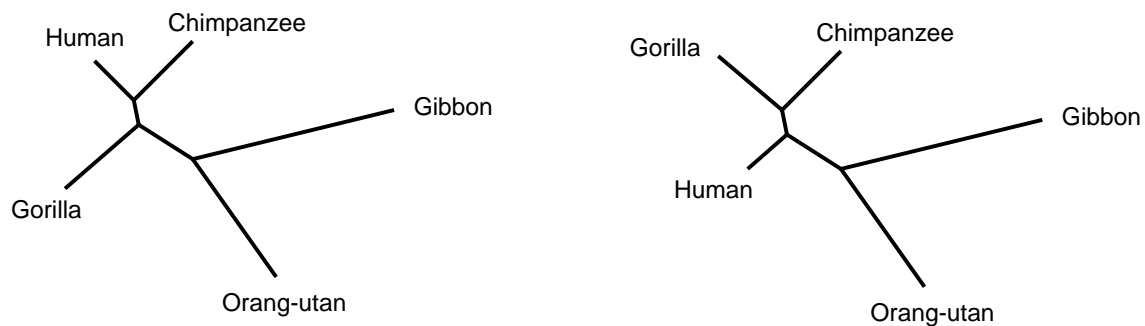
$$\begin{aligned} L_A &= \ln P(D|S_A, \hat{\mathbf{w}}_A) \\ L_B &= \ln P(D|S_B, \hat{\mathbf{w}}_B) \end{aligned} \longrightarrow \Delta L = L_A - L_B > 0$$

Is the difference in  $L$  between the two trees significant?

---

## Nonparametric bootstrapping

---



$$\begin{aligned} L_A &= \ln P(D|S_A, \hat{\mathbf{w}}_A) \\ L_B &= \ln P(D|S_B, \hat{\mathbf{w}}_B) \end{aligned} \longrightarrow \Delta L = L_A - L_B \neq 0$$

Resample the columns  $\mathbf{y}_t$  with replacement from the alignment  $D$ :

$$\begin{aligned} D &= \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4, \dots\} \longrightarrow \\ &D_1 = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_2, \mathbf{y}_4, \dots\} \\ &D_2 = \{\mathbf{y}_2, \mathbf{y}_2, \mathbf{y}_1, \mathbf{y}_1, \dots\} \\ &\vdots \\ &D_B = \{\mathbf{y}_3, \mathbf{y}_4, \mathbf{y}_4, \mathbf{y}_1, \dots\} \end{aligned}$$

## Nonparametric bootstrapping



$$\begin{aligned}
 L_A &= \ln P(D|S_A, \hat{\mathbf{w}}_A) \\
 L_B &= \ln P(D|S_B, \hat{\mathbf{w}}_B)
 \end{aligned}
 \longrightarrow \Delta L = L_A - L_B \neq 0$$

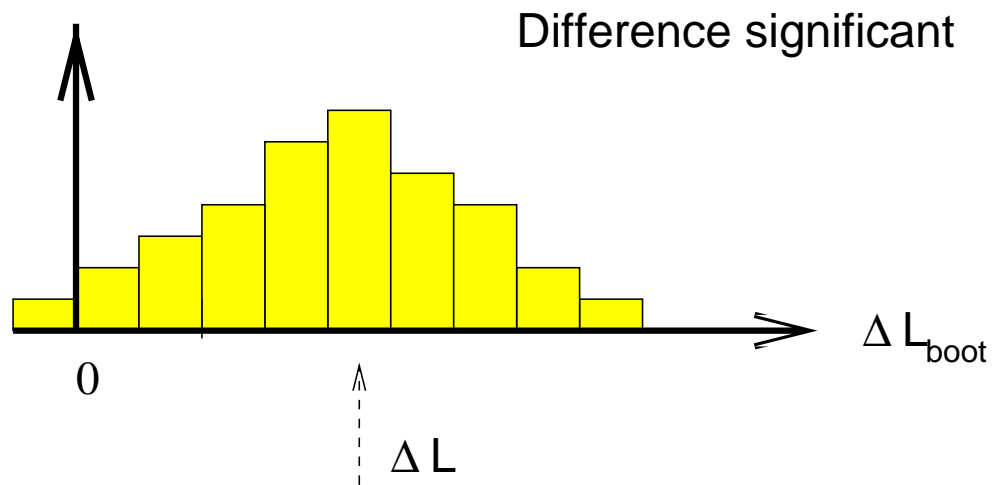
Resample the columns  $\mathbf{y}_t$  with replacement from the alignment D:

$$\begin{aligned}
 D &= \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4, \dots\} \longrightarrow \\
 &D_1 = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_2, \mathbf{y}_4, \dots\} \\
 &D_2 = \{\mathbf{y}_2, \mathbf{y}_2, \mathbf{y}_1, \mathbf{y}_1, \dots\} \\
 &\vdots \\
 &D_B = \{\mathbf{y}_3, \mathbf{y}_4, \mathbf{y}_4, \mathbf{y}_1, \dots\}
 \end{aligned}$$

Repeat the analysis on  $\{D_1, \dots, D_B\} \rightarrow$  Bootstrap distribution  $\{\Delta L_b\}_{b=1}^B$

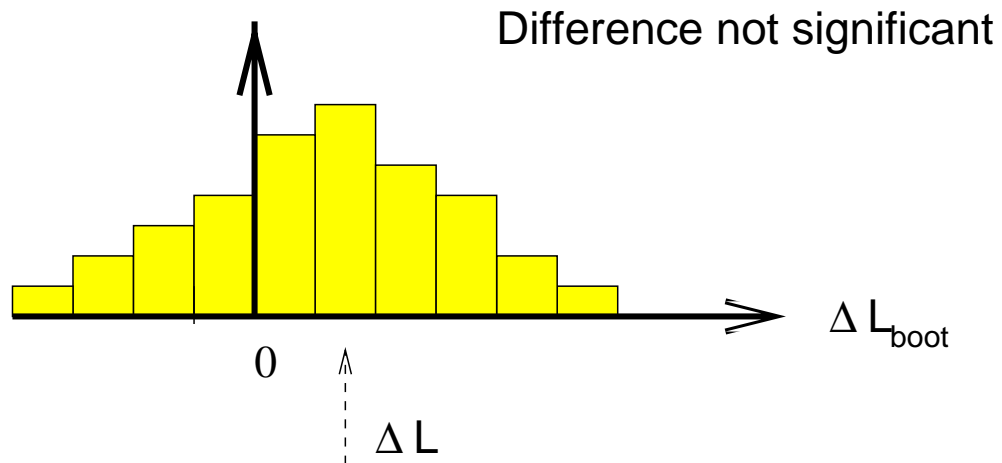
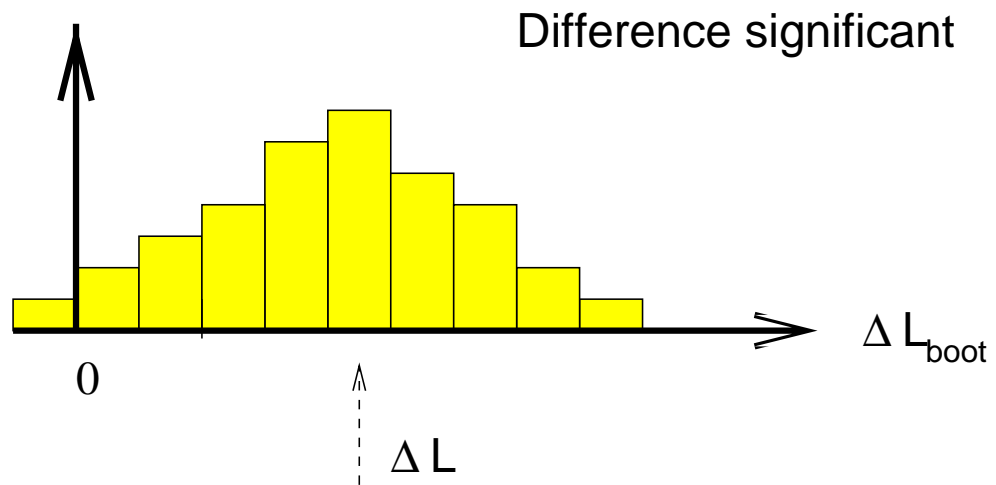
## Nonparametric bootstrapping

---



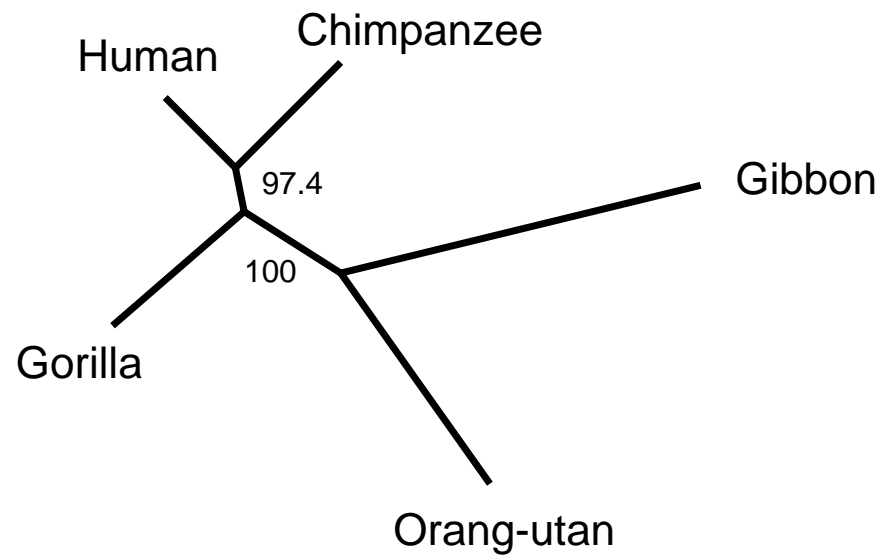
## Nonparametric bootstrapping

---



## Bootstrap support of clade formations

---



Clade	Probability
( Human Chimp )	0.974
( Human Chimp Gorilla )	1.0

---

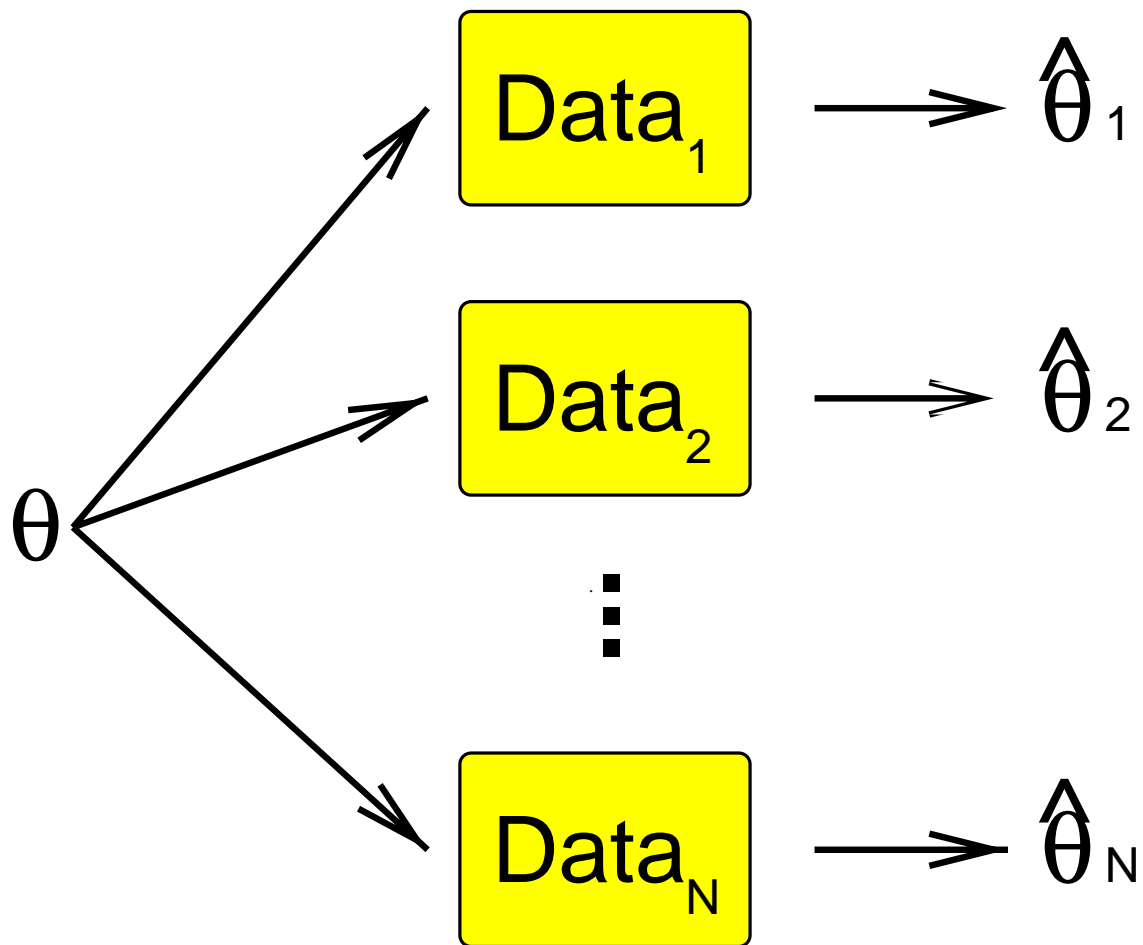
Statistical inference

---



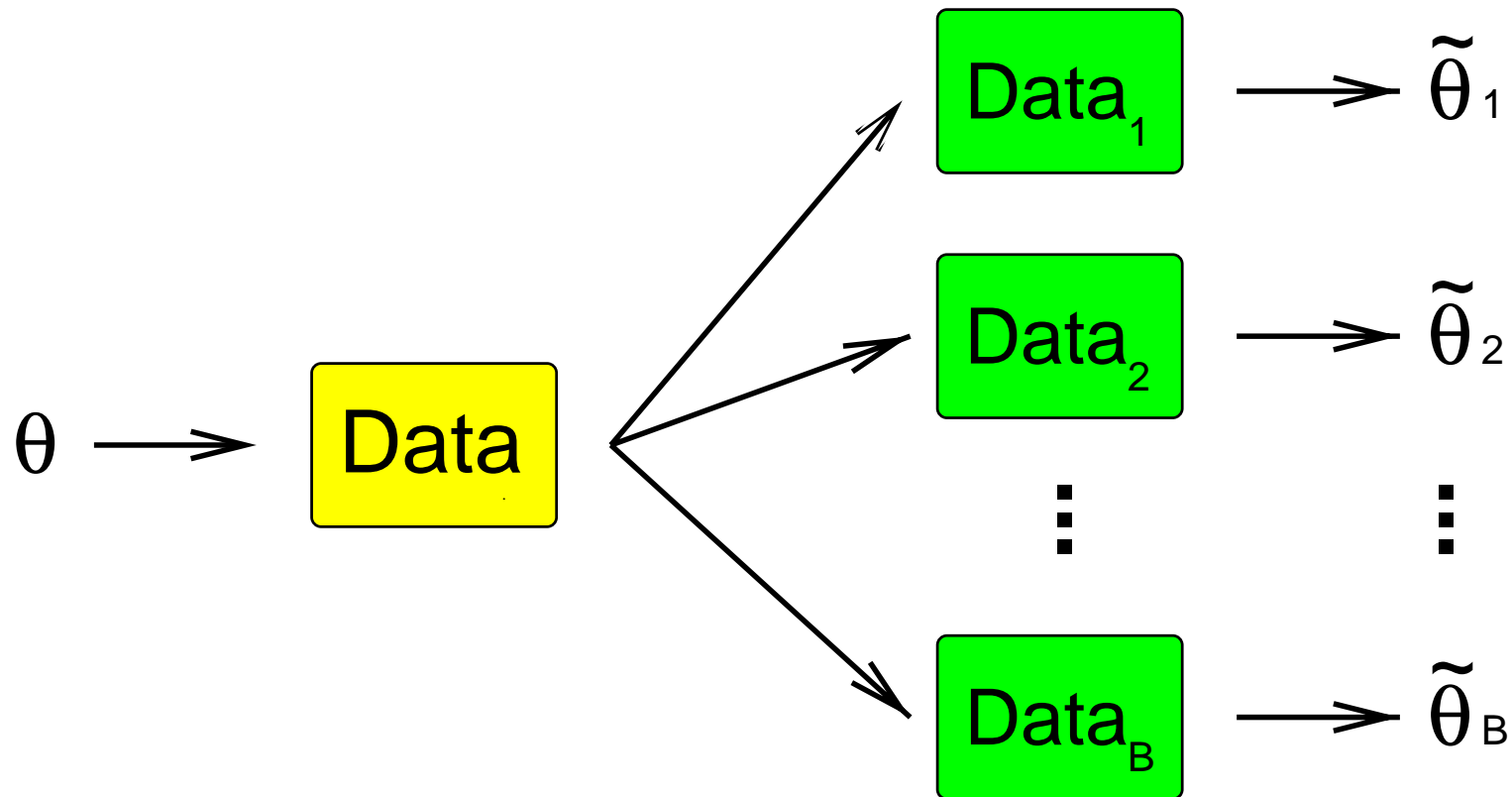
## Statistical inference

---



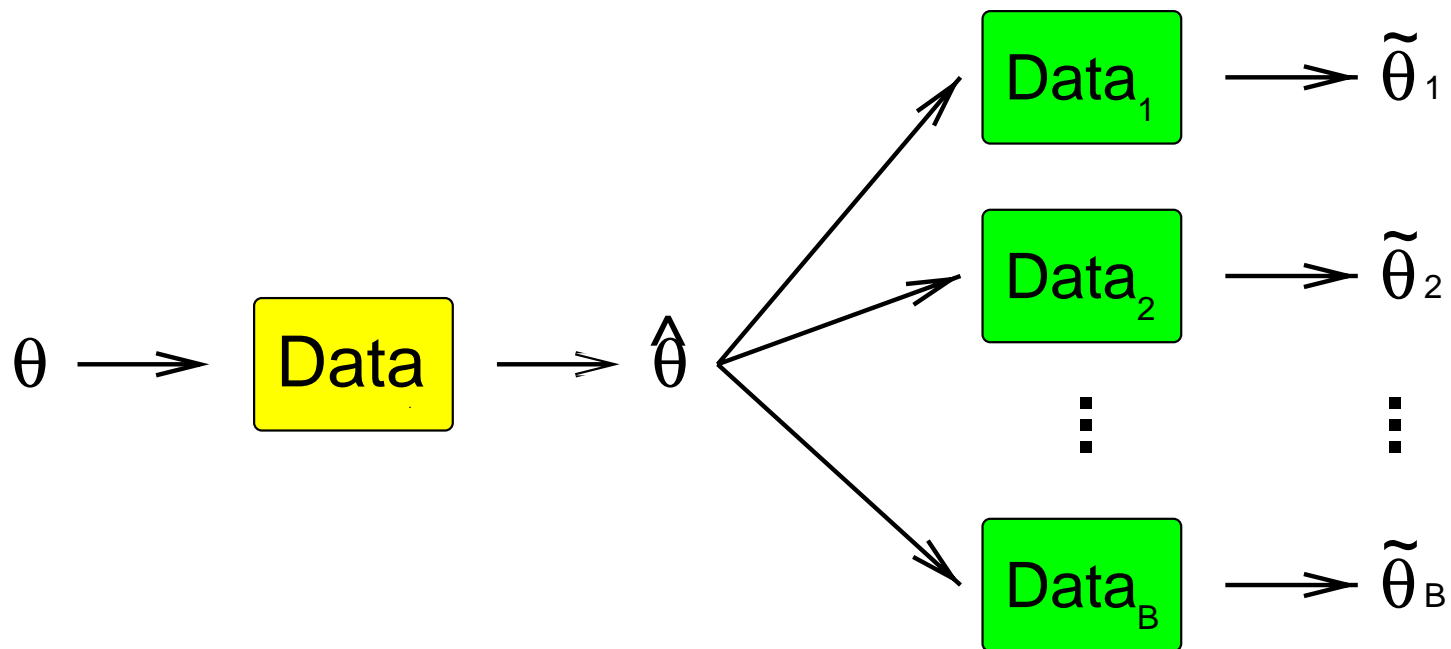
## Frequentist statistics: Nonparametric bootstrapping

---



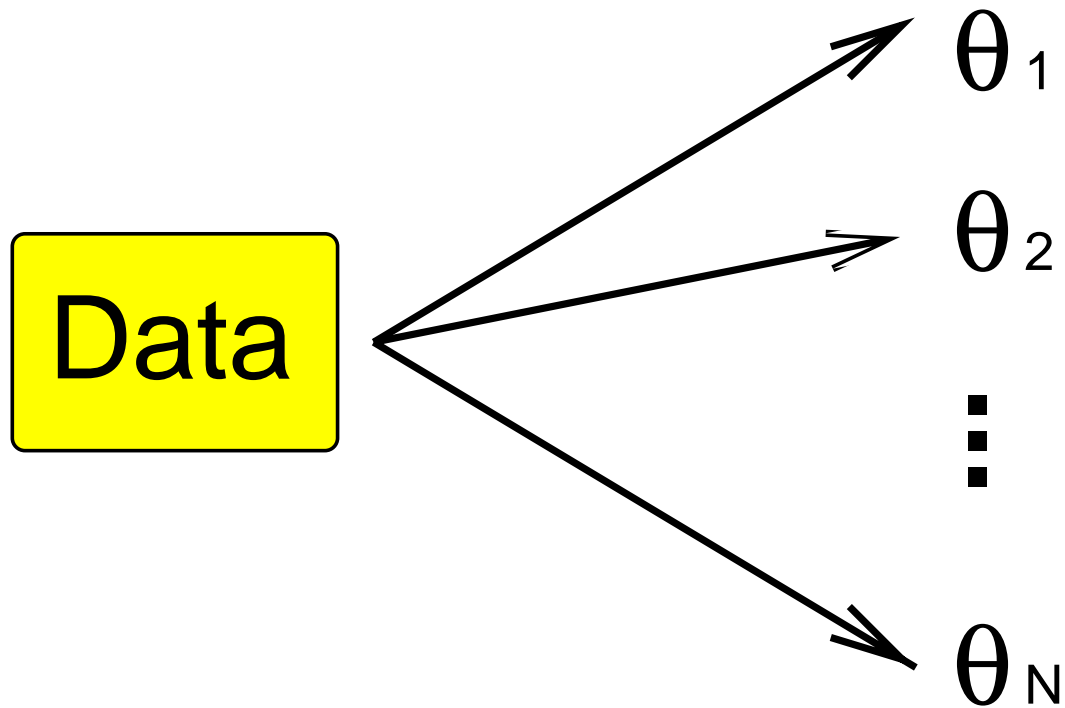
## Frequentist statistics: Parametric Bootstrapping

---



## Bayesian statistics

---



Posterior probability:  $P(\theta|\text{Data})$

---

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
-

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
  - Devise a Markov chain  $P^{(n+1)}(\theta_i) = \sum_k T(\theta_i|\theta_k)P^{(n)}(\theta_k)$  that converges in distribution to  $P(\theta|D)$ :  $P^{(n)}(\theta) \rightarrow P(\theta|D)$
-

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
  - Devise a Markov chain  $P^{(n+1)}(\theta_i) = \sum_k T(\theta_i|\theta_k)P^{(n)}(\theta_k)$  that converges in distribution to  $P(\theta|D)$ :  $P^{(n)}(\theta) \rightarrow P(\theta|D)$
  - Theorem: An **ergodic** Markov chain converges to its **stationary distribution** irrespective of its **initialization** .
-

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
  - Devise a Markov chain  $P^{(n+1)}(\theta_i) = \sum_k T(\theta_i|\theta_k)P^{(n)}(\theta_k)$  that converges in distribution to  $P(\theta|D)$ :  $P^{(n)}(\theta) \rightarrow P(\theta|D)$
  - Theorem: An **ergodic** Markov chain converges to its **stationary distribution** irrespective of its **initialization** .
  - Stationary distribution:  $P(\theta_i) = \sum_k T(\theta_i|\theta_k)P(\theta_k)$
-

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
  - Devise a Markov chain  $P^{(n+1)}(\theta_i) = \sum_k T(\theta_i|\theta_k)P^{(n)}(\theta_k)$  that converges in distribution to  $P(\theta|D)$ :  $P^{(n)}(\theta) \rightarrow P(\theta|D)$
  - Theorem: An **ergodic** Markov chain converges to its **stationary distribution** irrespective of its **initialization** .
  - Stationary distribution:  $P(\theta_i) = \sum_k T(\theta_i|\theta_k)P(\theta_k)$
  - Design the **Markov transition matrix**  $T(\theta_i|\theta_k)$  such that  $P(\theta|D)$  is the stationary distribution.
-

## Metropolis-Hastings algorithm

---

- Objective: Sample from the **posterior distribution**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- Direct approach intractable due to  $\int P(D|\theta)P(\theta)d\theta$
  - Devise a Markov chain  $P^{(n+1)}(\theta_i) = \sum_k T(\theta_i|\theta_k)P^{(n)}(\theta_k)$  that converges in distribution to  $P(\theta|D)$ :  $P^{(n)}(\theta) \rightarrow P(\theta|D)$
  - Theorem: An **ergodic** Markov chain converges to its **stationary distribution** irrespective of its **initialization** .
  - Stationary distribution:  $P(\theta_i) = \sum_k T(\theta_i|\theta_k)P(\theta_k)$
  - Design the **Markov transition matrix**  $T(\theta_i|\theta_k)$  such that  $P(\theta|D)$  is the stationary distribution.
  - Detailed balance:  $\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(\theta_k|D)}{P(\theta_i|D)} = \frac{P(D|\theta_k)P(\theta_k)}{P(D|\theta_i)P(\theta_i)}$
-

## Proof

---

Show that  $\sum_k T(\theta_i|\theta_k)P(\theta_k|D) = P(\theta_i|D)$

---

## Proof

---

Show that  $\sum_k T(\theta_i|\theta_k)P(\theta_k|D) = P(\theta_i|D)$

Detailed balance:  $\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(\theta_k|D)}{P(\theta_i|D)} \implies$

---

## Proof

---

Show that  $\sum_k T(\theta_i|\theta_k)P(\theta_k|D) = P(\theta_i|D)$

Detailed balance:  $\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(\theta_k|D)}{P(\theta_i|D)} \implies$

$$\begin{aligned} \sum_k T(\theta_i|\theta_k)P(\theta_k|D) &= \sum_k T(\theta_k|\theta_i)P(\theta_i|D) = \\ &P(\theta_i|D) \sum_k T(\theta_k|\theta_i) = P(\theta_i|D) \end{aligned}$$

---

## Metropolis-Hastings algorithm

---

$$\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(D|\theta_k)P(\theta_k)}{P(D|\theta_i)P(\theta_i)}$$

Transition Probability = Proposal Probability  $\times$  Acceptance Probability

$$T(\theta_k|\theta_i) = q(\theta_k|\theta_i)a(\theta_k|\theta_i)$$

---

## Metropolis-Hastings algorithm

---

$$\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(D|\theta_k)P(\theta_k)}{P(D|\theta_i)P(\theta_i)}$$

Transition Probability = Proposal Probability  $\times$  Acceptance Probability

$$T(\theta_k|\theta_i) = q(\theta_k|\theta_i)a(\theta_k|\theta_i)$$

Acceptance Probabilities:

$$\frac{a(\theta_k|\theta_i)}{a(\theta_i|\theta_k)} = \frac{P(D|\theta_k)P(\theta_k)q(\theta_i|\theta_k)}{P(D|\theta_i)P(\theta_i)q(\theta_k|\theta_i)}$$

---

## Metropolis-Hastings algorithm

---

$$\frac{T(\theta_k|\theta_i)}{T(\theta_i|\theta_k)} = \frac{P(D|\theta_k)P(\theta_k)}{P(D|\theta_i)P(\theta_i)}$$

Transition Probability = Proposal Probability  $\times$  Acceptance Probability

$$T(\theta_k|\theta_i) = q(\theta_k|\theta_i)a(\theta_k|\theta_i)$$

Acceptance Probabilities:

$$\frac{a(\theta_k|\theta_i)}{a(\theta_i|\theta_k)} = \frac{P(D|\theta_k)P(\theta_k)q(\theta_i|\theta_k)}{P(D|\theta_i)P(\theta_i)q(\theta_k|\theta_i)}$$

$$a(\theta_k|\theta_i) = \min \left\{ \frac{P(D|\theta_k)P(\theta_k)q(\theta_i|\theta_k)}{P(D|\theta_i)P(\theta_i)q(\theta_k|\theta_i)}, 1 \right\}$$

---

## Metropolis-Hastings algorithm

---

- Start from initial  $\theta_0$

## Metropolis-Hastings algorithm

---

- Start from initial  $\theta_0$
- Iterate  $n = 1 \dots N$ 
  1. Obtain new  $\theta^{(n)}$  from proposal distribution  $q(\theta^{(n)} | \theta^{(n-1)})$

## Metropolis-Hastings algorithm

---

- Start from initial  $\theta_0$
- Iterate  $n = 1 \dots N$ 
  1. Obtain new  $\theta^{(n)}$  from proposal distribution  $q(\theta^{(n)}|\theta^{(n-1)})$
  2. Accept with probability  $a(\theta^{(n)}|\theta^{(n-1)})$ , otherwise leave unchanged:  $\theta^{(n)} = \theta^{(n-1)}$

## Metropolis-Hastings algorithm

---

- Start from initial  $\theta_0$
  - Iterate  $n = 1 \dots N$ 
    1. Obtain new  $\theta^{(n)}$  from proposal distribution  $q(\theta^{(n)}|\theta^{(n-1)})$
    2. Accept with probability  $a(\theta^{(n)}|\theta^{(n-1)})$ , otherwise leave unchanged:  $\theta^{(n)} = \theta^{(n-1)}$
  - Discard  $\theta_1, \dots, \theta_{N/2}$  (burn-in period)
-

## Metropolis-Hastings algorithm

---

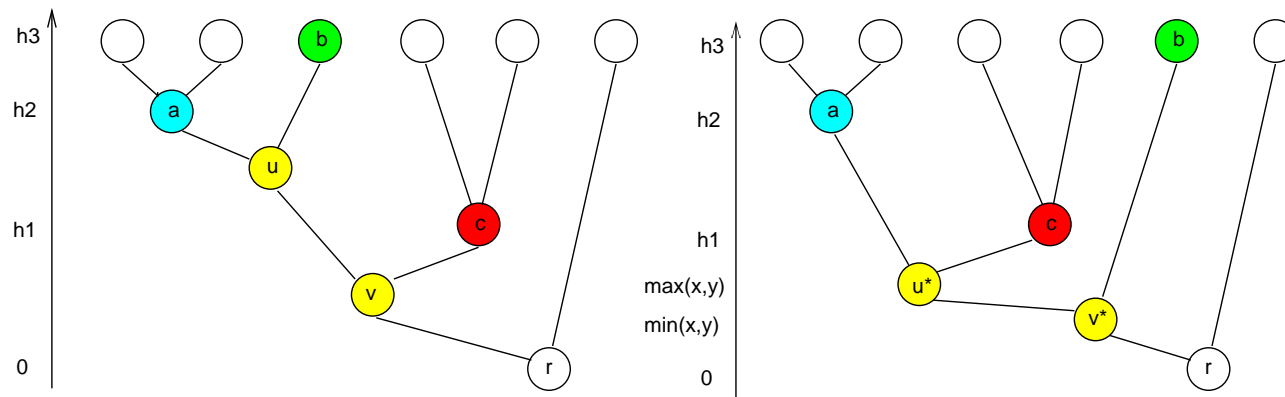
- Start from initial  $\theta_0$
  - Iterate  $n = 1 \dots N$ 
    1. Obtain new  $\theta^{(n)}$  from proposal distribution  $q(\theta^{(n)}|\theta^{(n-1)})$
    2. Accept with probability  $a(\theta^{(n)}|\theta^{(n-1)})$ , otherwise leave unchanged:  $\theta^{(n)} = \theta^{(n-1)}$
  - Discard  $\theta_1, \dots, \theta_{N/2}$  (burn-in period)
  - Sample from  $\theta_{N/2+1}, \dots, \theta_N$
-

## Metropolis-Hastings algorithm

---

- Start from initial  $\theta_0$
  - Iterate  $n = 1 \dots N$ 
    1. Obtain new  $\theta^{(n)}$  from proposal distribution  $q(\theta^{(n)}|\theta^{(n-1)})$
    2. Accept with probability  $a(\theta^{(n)}|\theta^{(n-1)})$ , otherwise leave unchanged:  $\theta^{(n)} = \theta^{(n-1)}$
  - Discard  $\theta_1, \dots, \theta_{N/2}$  (burn-in period)
  - Sample from  $\theta_{N/2+1}, \dots, \theta_N$
  - $$\int f(\theta)P(\theta|D)d\theta = \frac{2}{N} \sum_{n=N/2+1}^N f(\theta_n)$$
-

## Moves in tree space and Hastings ratio

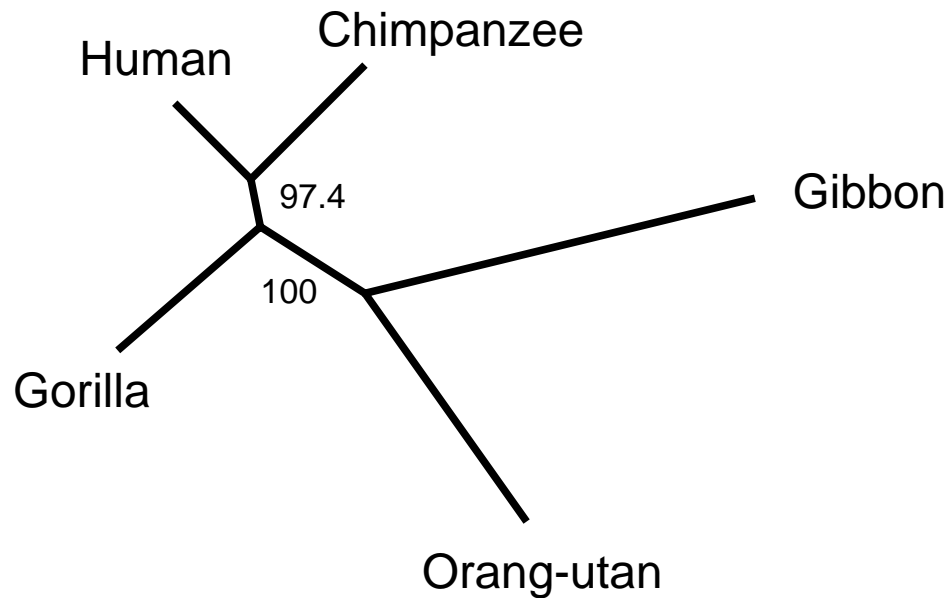


- $x$  uniformly chosen at random from  $[0, h_1]$ ,  $y$  uniformly chosen at random from  $[0, h_2]$
- New nodes  $u^*$  and  $v^*$ :  $\text{distance}(r, u^*) = \max(x, y)$      $\text{distance}(r, v^*) = \min(x, y)$
- $\max(x, y) > h_1 \implies$  Leave topology unchanged
- $\max(x, y) < h_1 \implies$  Choose between three possible topologies

Current tree	$\max(x, y)$	Hastings ratio
$\text{distance}(v, u) > \text{distance}(v, c)$	$< h_1$	3
$\text{distance}(v, u) < \text{distance}(v, c)$	$> h_1$	1/3
otherwise		1

## Posterior probability of clades

---



Clade	Probability
( Human Chimp )	0.974
( Human Chimp Gorilla )	1.0

---