Tree Dating using Beast2

Demo Day 3, IPS-164 Introduction to Phylogenetics

Install Software

- Beast2 https://www.beast2.org
- Tracer http://tree.bio.ed.ac.uk/software/tracer/
- Fig tree http://tree.bio.ed.ac.uk/software/figtree/

- We are going to use this Tutorial Divergence Dating (Version 2.2.0) (https://www.beast2.org/tutorials/)

Main steps of tree dating analyses

- Getting orthologous sequences
- Alignment
- Selecting best partitioning scheme and substitution models

Main steps of tree dating analyses

- Getting orthologous sequences
- Alignment
- Selecting best partitioning scheme and substitution models
- Setting up priors for the substitution models

Bayesian theorem applied to probability distribution

Likelihood:
Probability of *B* given *A*

Prior:
Probability of *A* before gathering the data

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Posterior:
Probability that *A* is true given *B* is observed

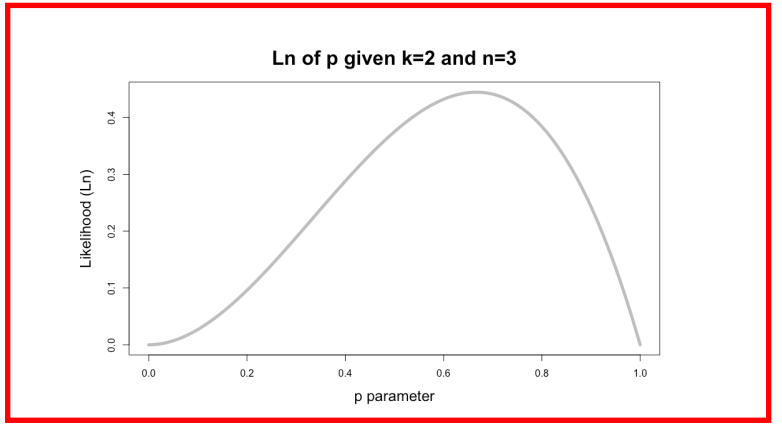
Probability of *B*
(=**probability of data**,
=**marginal probability**)

Without loss of generality posterior can be written as:

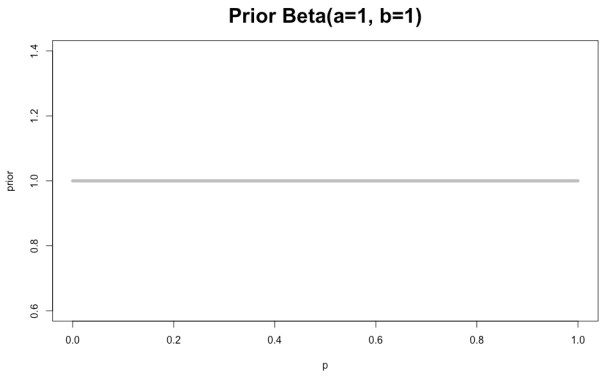
*Posterior ∝ Likelihood * Prior*

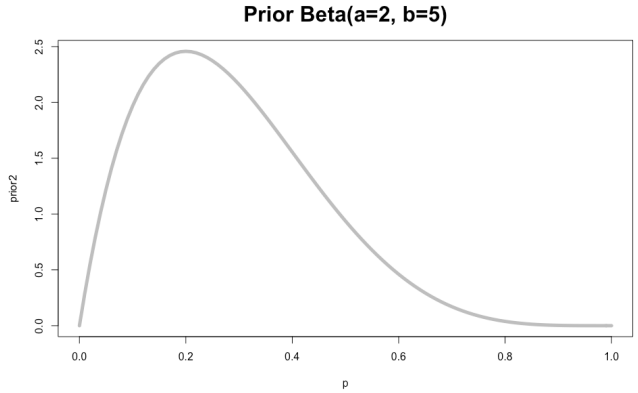
Posterior for coin toss

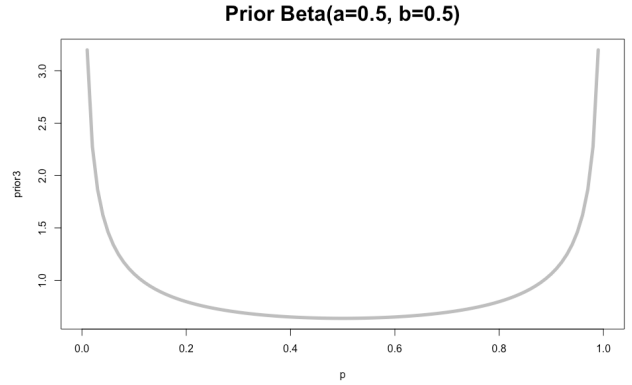
*Posterior ∝ Likelihood * Prior*



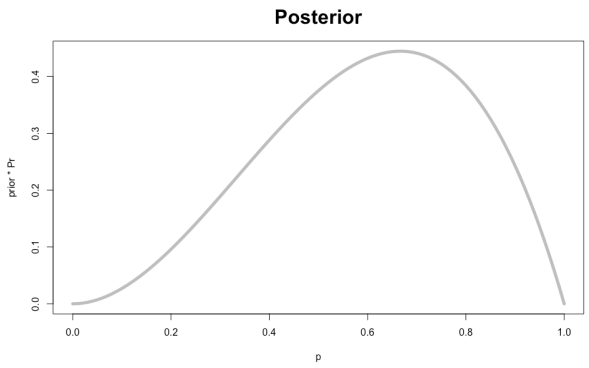
Priors:

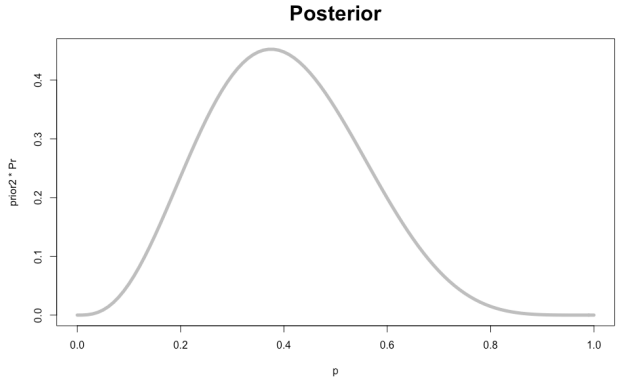


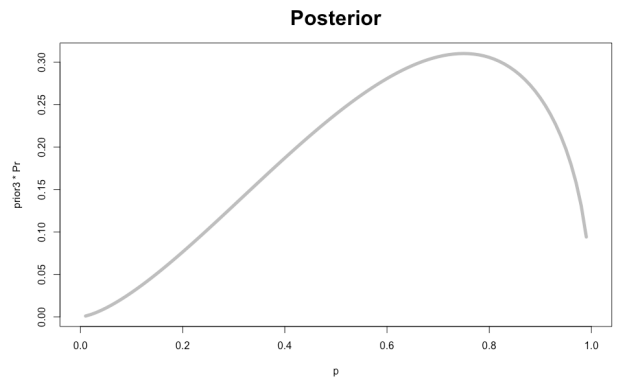




Posteriors:



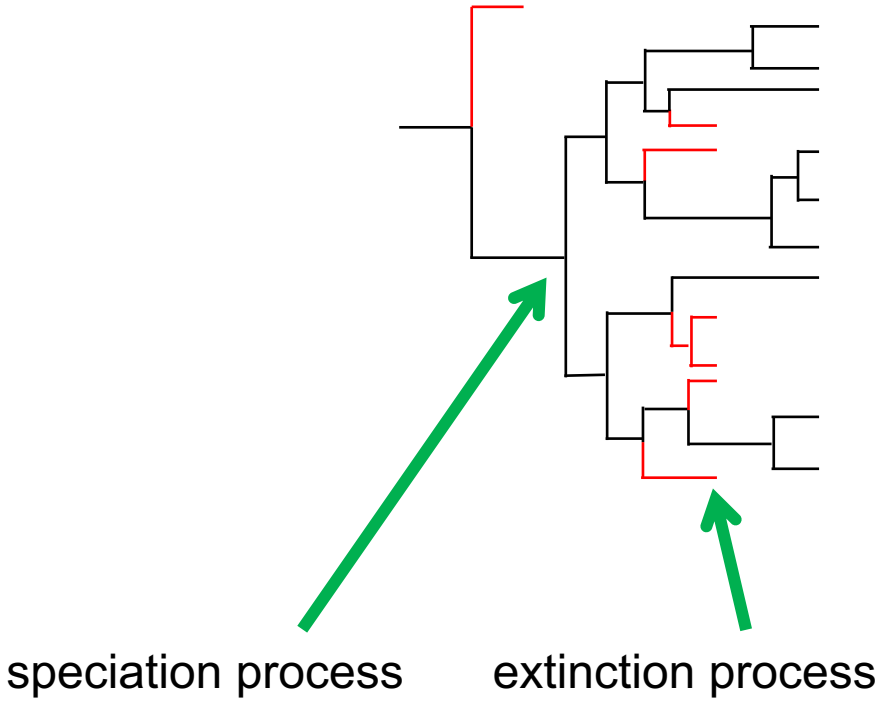




Main steps of tree dating analyses

- Getting orthologous sequences
- Alignment
- Selecting best partitioning scheme and substitution models
- Setting up priors for the substitution models
- Selecting clock and tree evolution models

Tree and clock model



Bayesian methods include
a prior on tree growth

Branching process prior

Pure birth

Birth-death

slide from Niklas Wahlberg

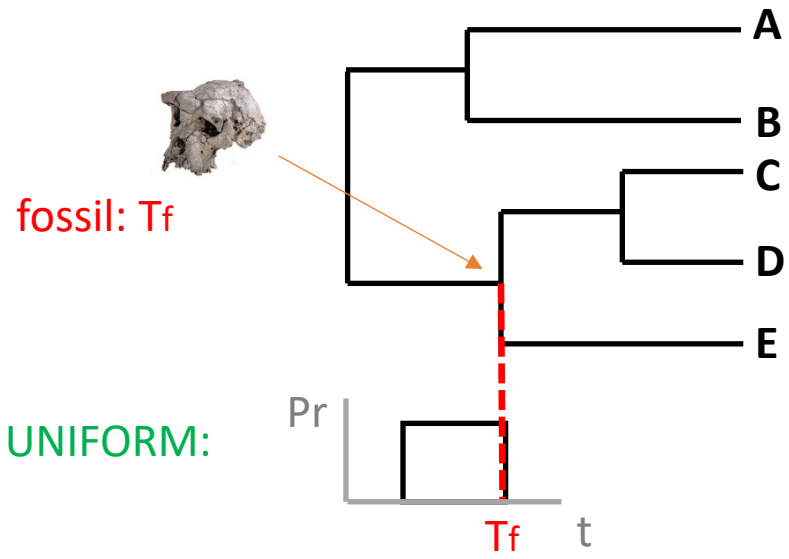
Main steps of tree dating analyses

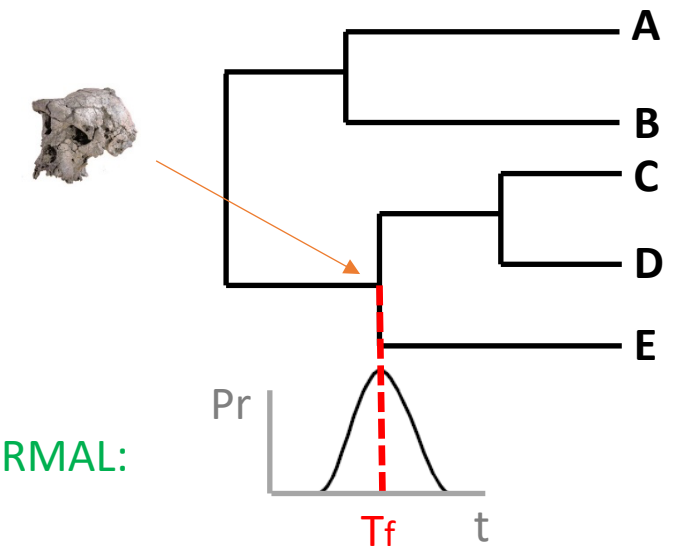
- Getting orthologous sequences
- Alignment
- Selecting best partitioning scheme and substitution models
- Setting up priors for the substitution models
- Selecting clock and tree evolution models
- Setting up priors for the clock and tree evolution models

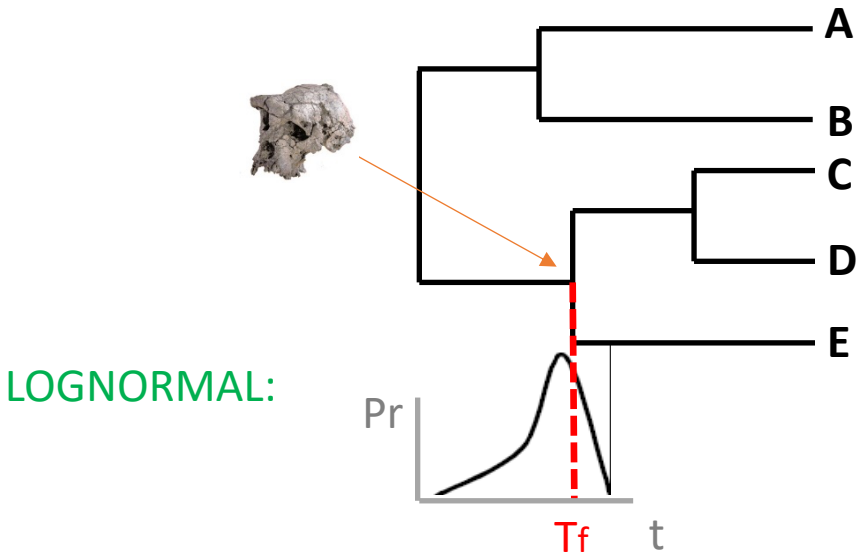
Main steps of tree dating analyses

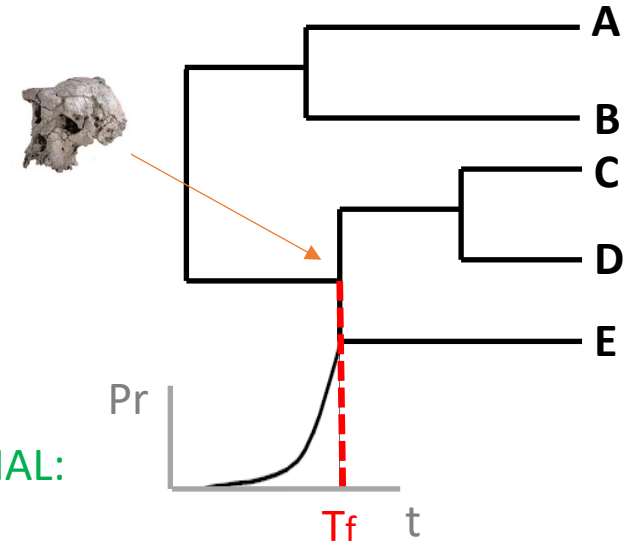
- Calibration: setting up priors for calibrating points

Prior distributions









slide from Niklas Wahlberg

Main steps of tree dating analyses

- Calibration: setting up priors for calibrating points
- Setting up MCMC

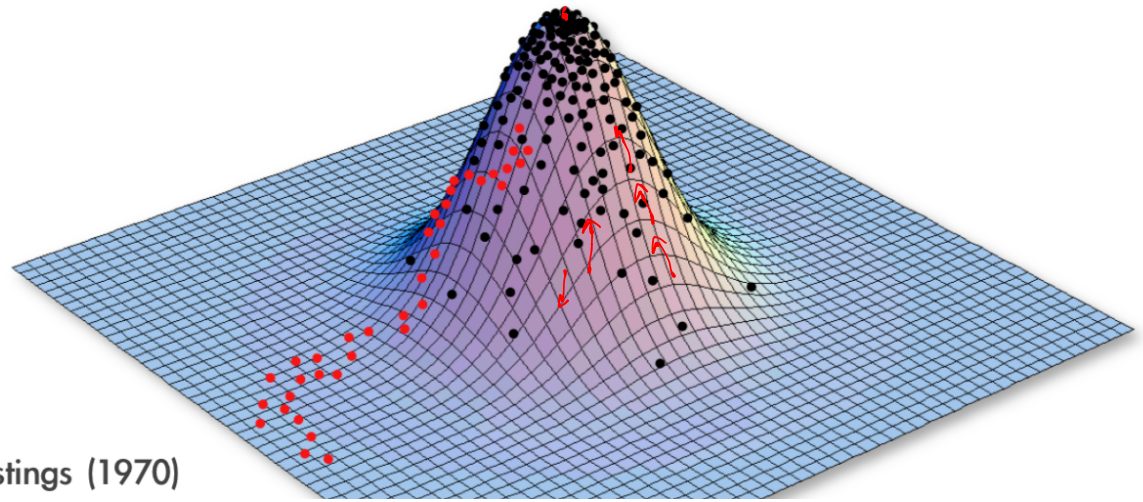
Approximating the Joint Posterior Probability Density with MCMC

Programming our MCMC robot...

Our robot parachutes into a random location in the joint posterior density and will explore parameter space by following these simple rules:

1. If the proposed step will take the robot uphill, it automatically takes the step
2. If the proposed step will take the robot downhill, it divides the elevation of the proposed location by the current location, and it only takes the step if the quotient is less than a uniform random variable, $U[0,1]$
3. The proposal distribution is symmetrical, so $Pr[A \rightarrow B] = Pr[B \rightarrow A]$

From the presentation of Brian Moore (Univ. of Davis)



Metropolis et al. (1953); Hastings (1970)

Main steps of tree dating analyses

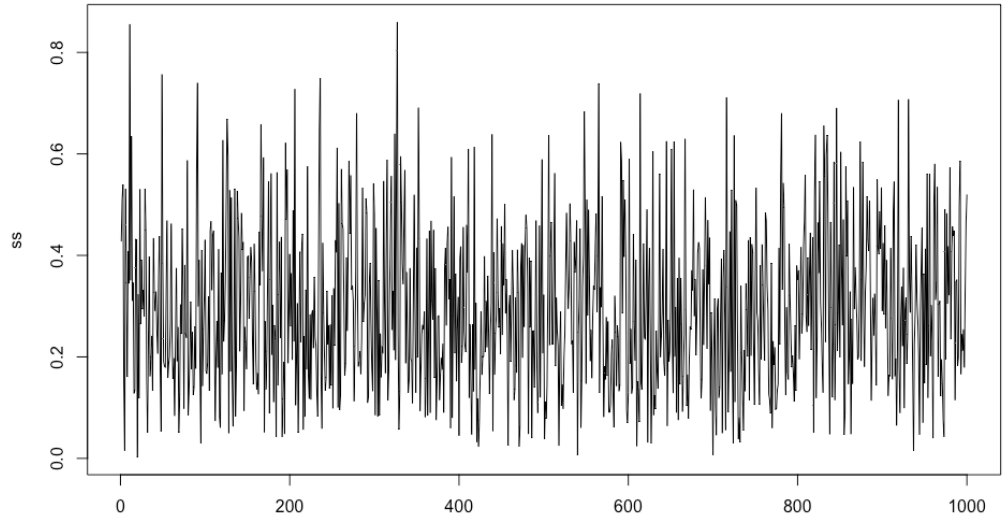
- Calibration: setting up priors for calibrating points
- Setting up MCMC
- Running analysis

Main steps of tree dating analyses

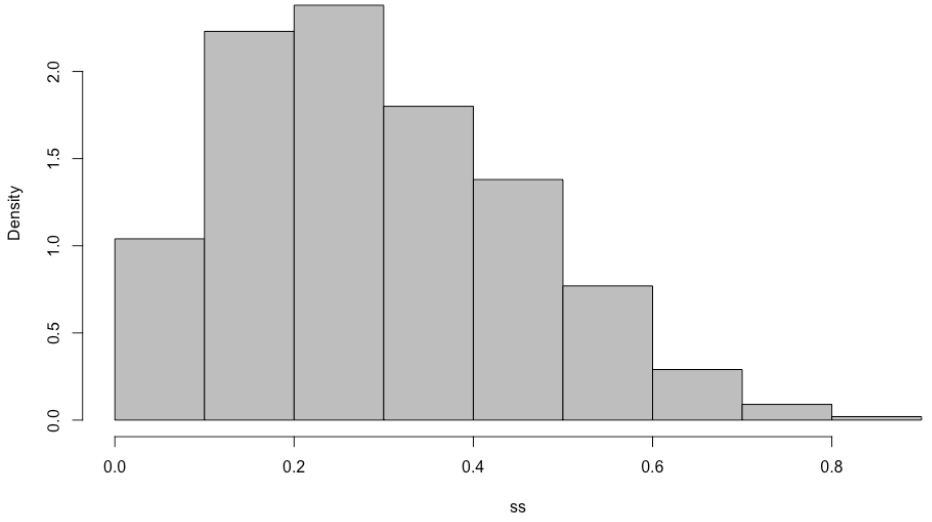
- Calibration: setting up priors for calibrating points
- Setting up MCMC
- Running analysis
- MCMC diagnostics

See Lecture 9 on MCMC diagnostics

MCMC is a Markov chain that being at stationary randomly samples from the posterior



Posterior



Main steps of tree dating analyses

- Calibration: setting up priors for calibrating points
- Setting up MCMC
- Running analysis
- MCMC diagnostics
- Interpreting your results