Model Selection

Laura Mulvey & Rachel Warnock

FAU Erlangen-Nürnberg

Wednesday 07.09.22

Model selections vs model adequacy

Model Selection and Testing

General Introduction to Model selection

Comparing relative model fit with Bayes factors

Model selection of common substitution models for one locus

Comparing relative model fit with Bayes factors

Model selection of partition models

Comparing relative model fit with Bayes factors

Model averaging of substitution models

Reversible-jump MCMC over substitution models

Introduction to Posterior Prediction

Assessing the fit of Normal distributions to trait data

**Assessing Phylogenetic Reliability Using RevBayes and P^3**

Model adequacy testing using posterior prediction (Data Version).

Assessing Phylogenetic Reliability Using RevBayes and P^3

Model adequacy testing using posterior prediction (Inference Version).

Revbayes tutorials

Impacts of substitution model on inferred parameters

Robinson-Foulds Distances

2_Demar
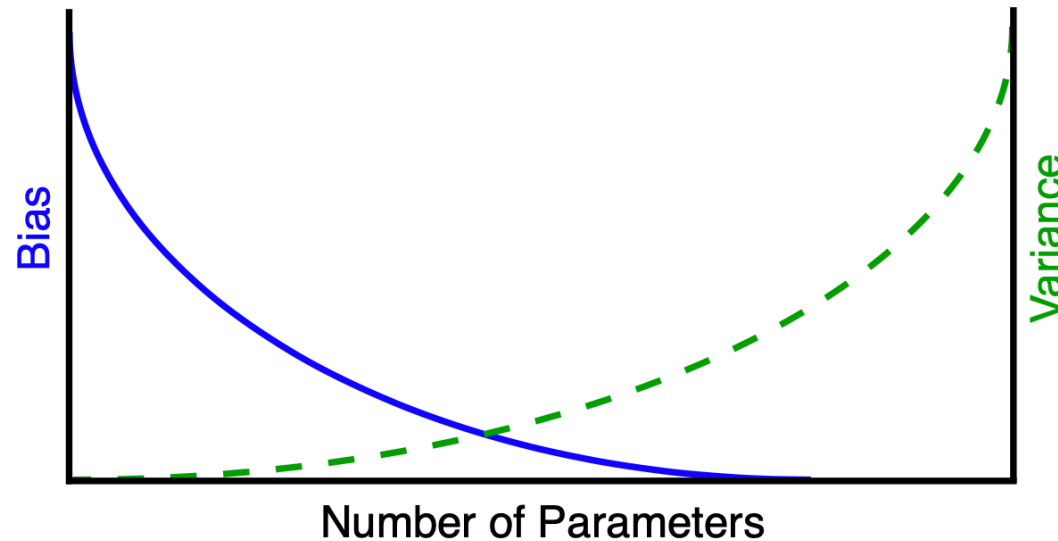Tree length range of posterior
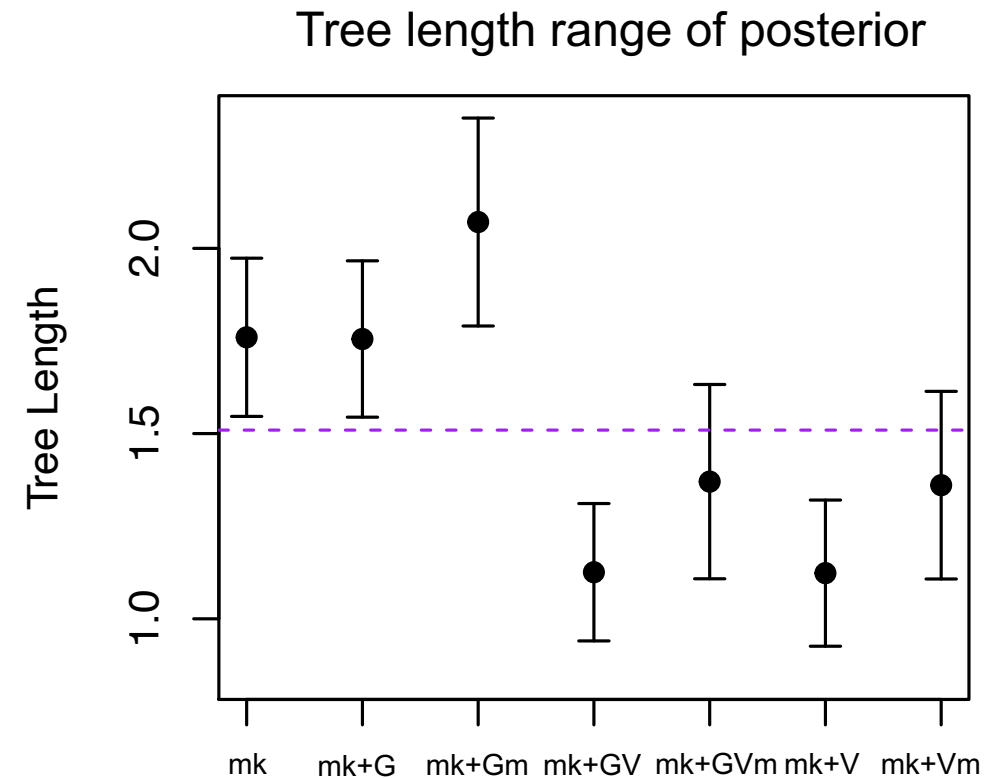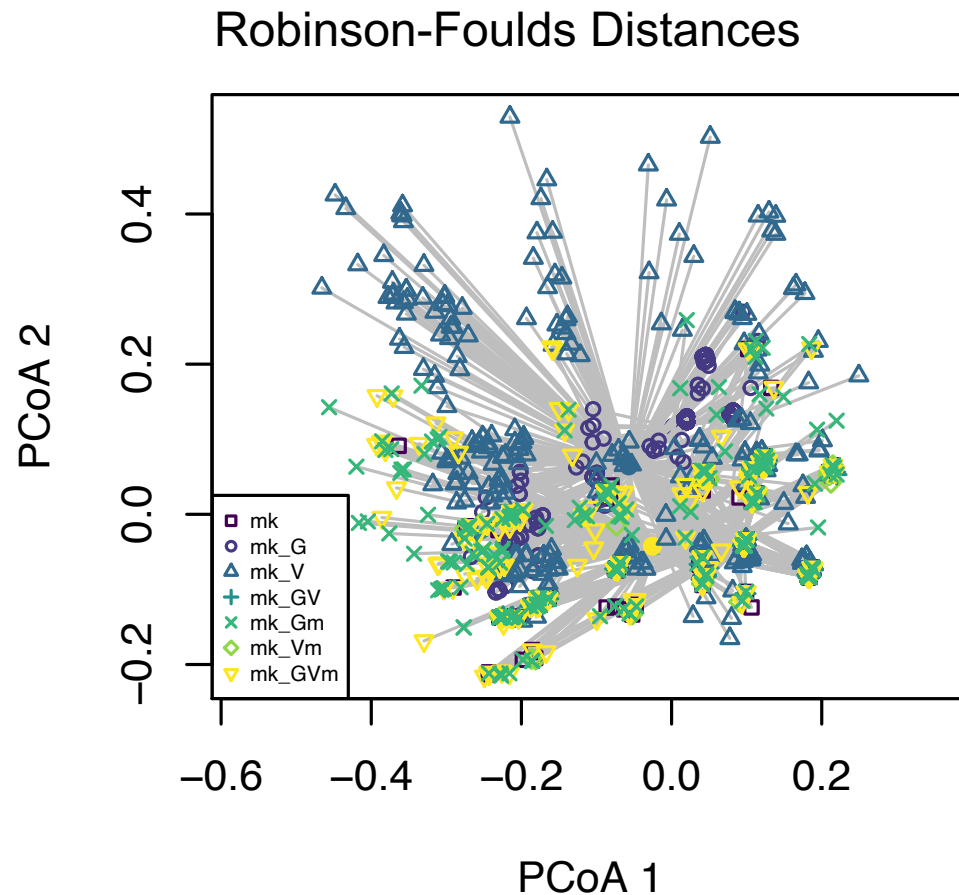
(Mulvey et al in prep.)

What does a good model look like?

To do statistical inference we need a model
What model should that be?
Our goal should be to have a model that is **complex enough** to capture the "important" variation in the data, but **not be more complex** than it needs to be

Too simple, misinterpreting the data

Too complicated, not enough information

Bias

Variance

Number of Parameters

What does a good model look like?

To do statistical inference we need a model
What model should that be?
Our goal should be to have a model that is **complex enough** to capture the "important" variation in the data, but **not be more complex** than it needs to be
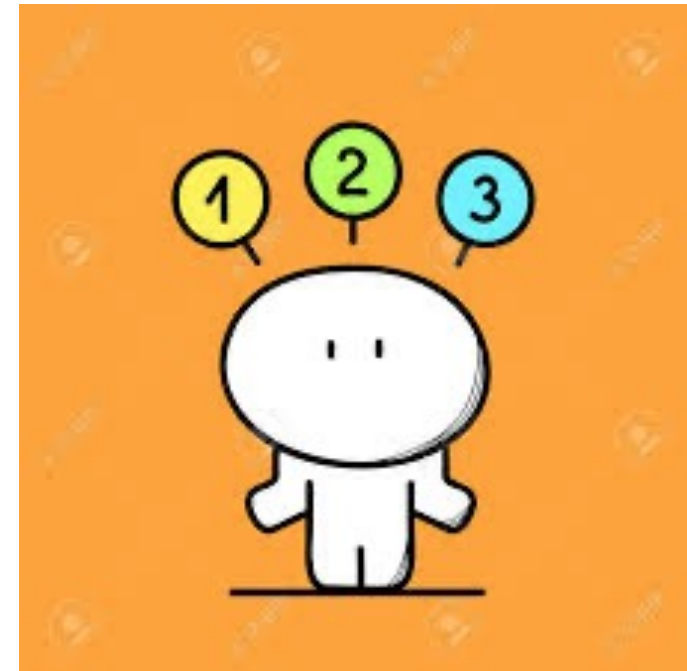

IT FITS PERFECT!
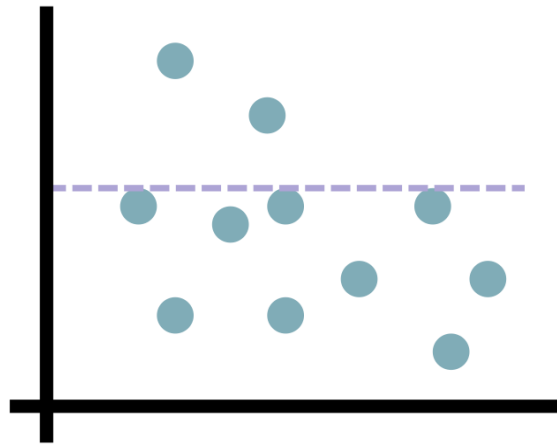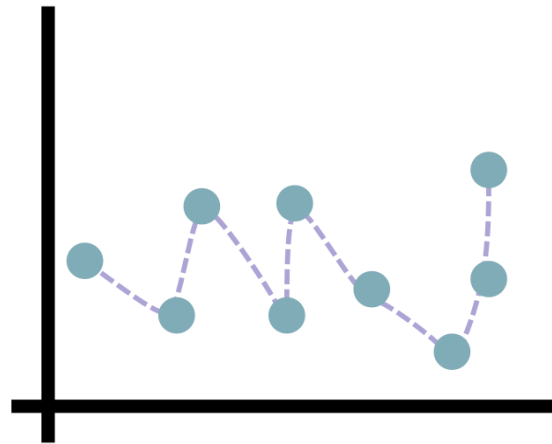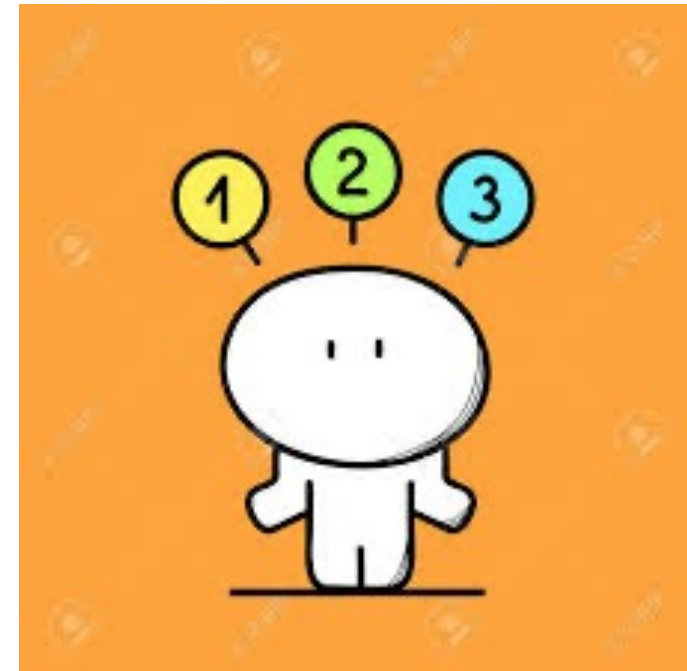

Underfitting

Overfitting
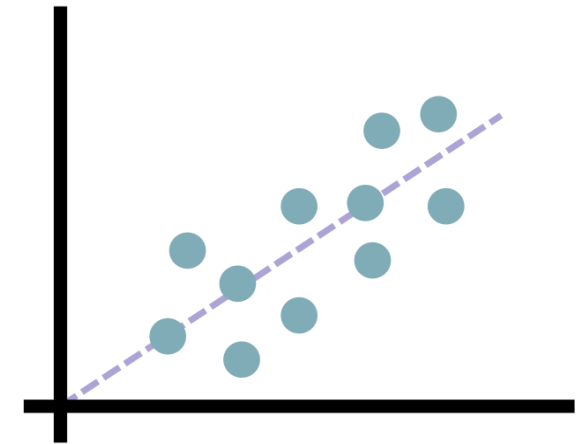
Proper fit

How to choose which model to use for morphological data?

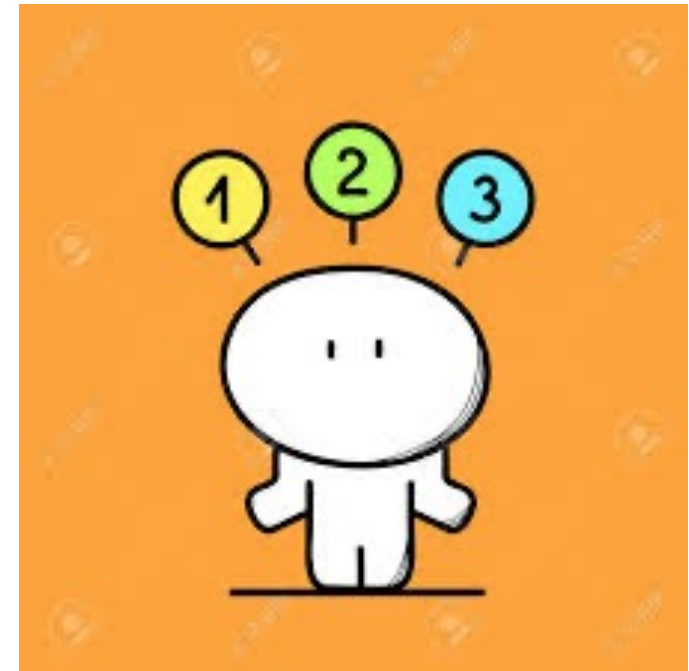# How to choose which model to use for morphological data?

Guess

How to choose which model to use for morphological data?
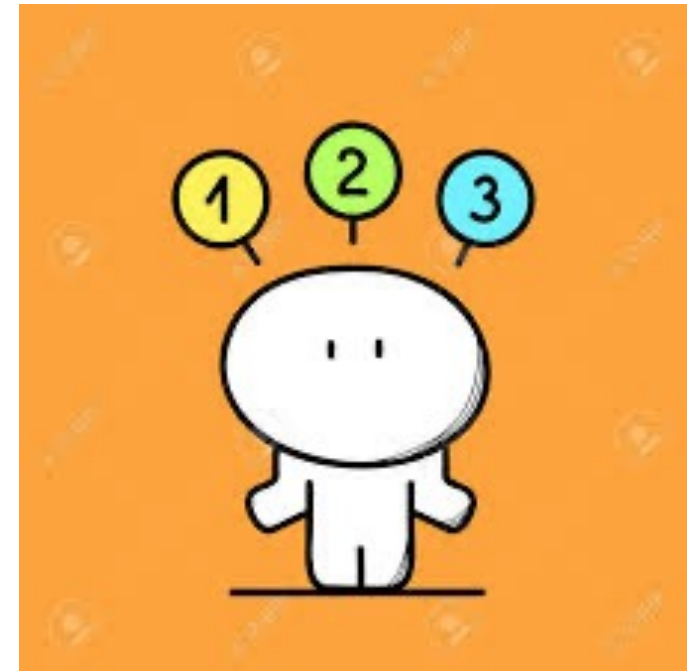
Guess
What other people have done

How to choose which model to use for morphological data?
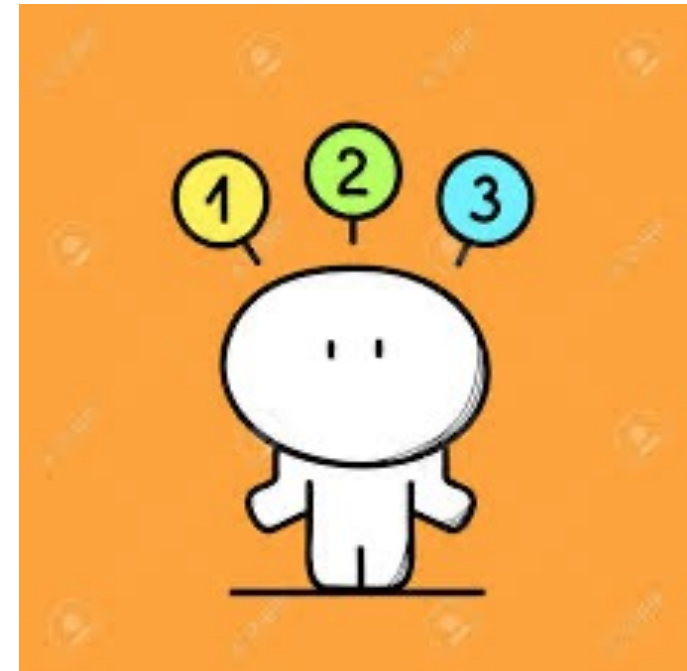
Guess
What other people have done
AIC values

How to choose which model to use for morphological data?

Guess
What other people have done
AIC values
BIC Values

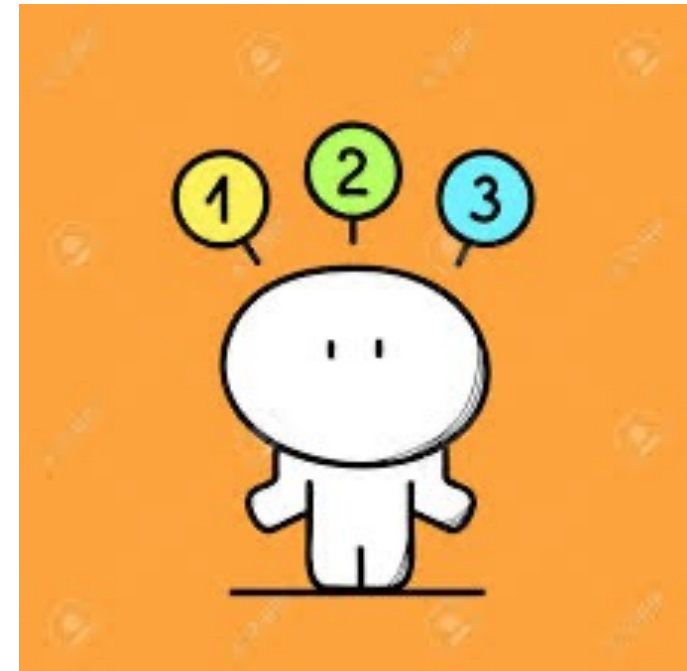# How to choose which model to use for morphological data?

Guess
What other people have done
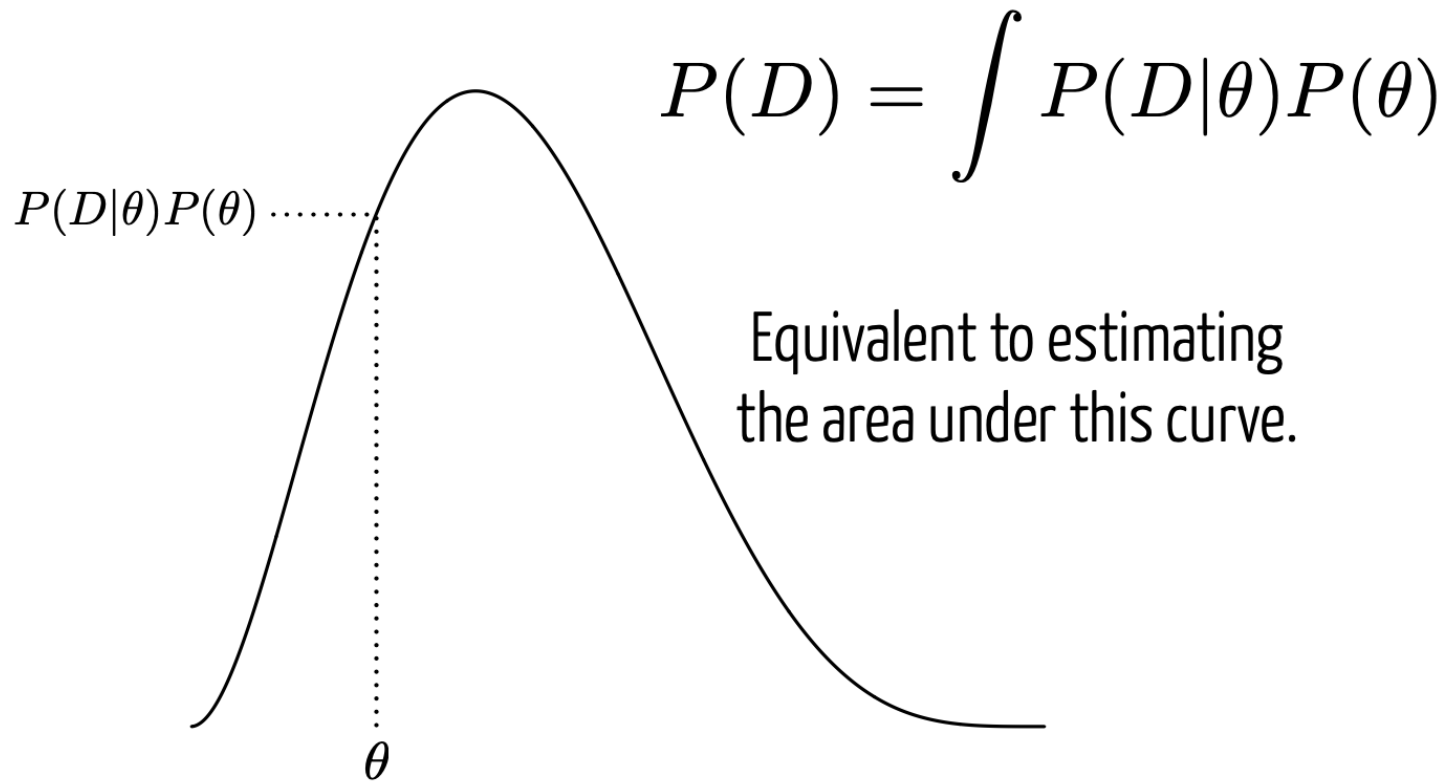AIC values
BIC Values
Bayes factors :

Bayes Theorem

P (parameters | data, model) =

Posterior

Likelihood

Priors

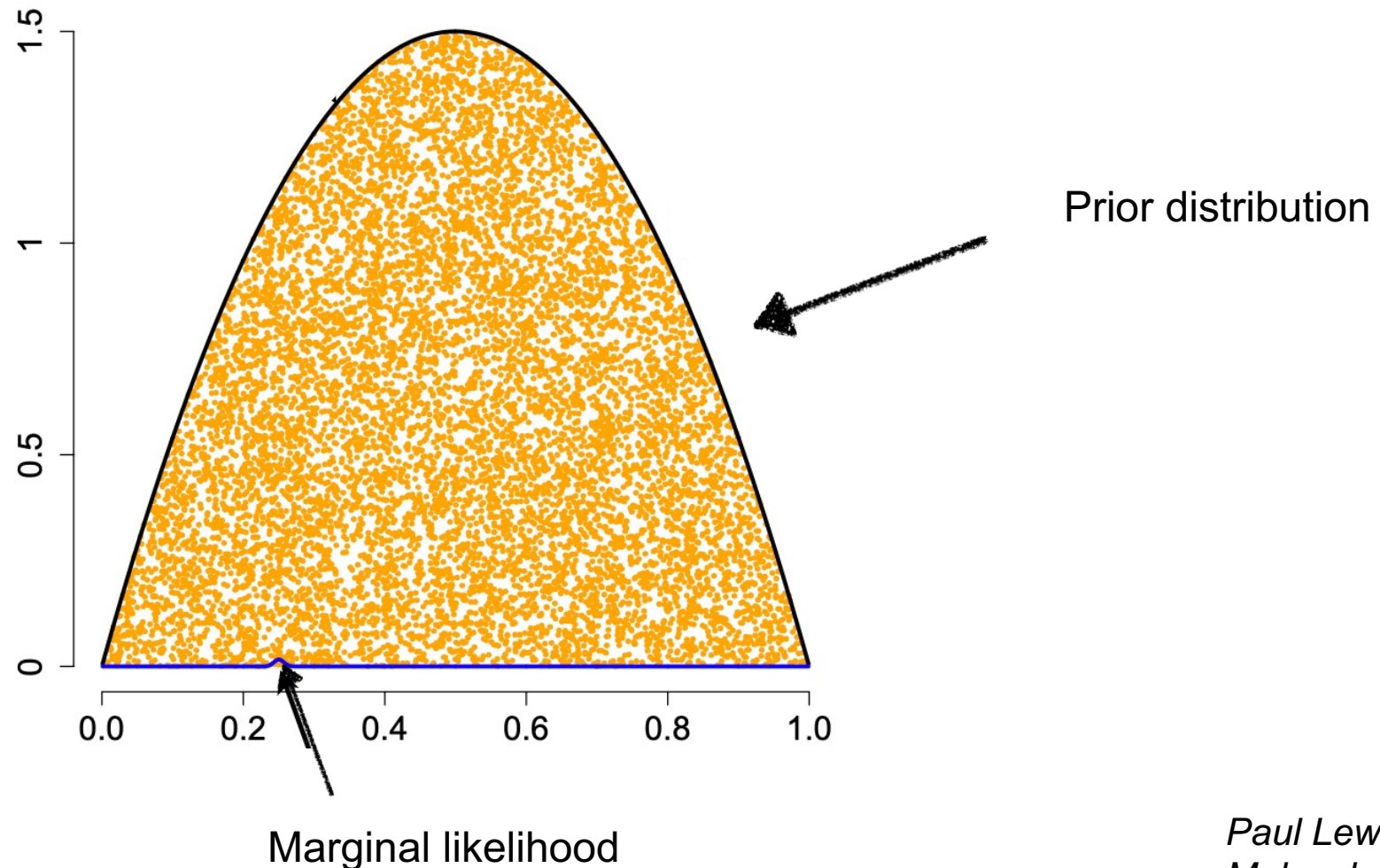P (data | parameters, model) P (parameters | model)

_____

P (data | model)

Marginal
probability

What is the marginal likelihood…..

$$P(D) = \int P(D|\theta)P(\theta)$$

$P(D|\theta)P(\theta)$ ·········

Equivalent to estimating the area under this curve.

θ

How can we estimate the marginal likelihood

Prior distribution

Marginal likelihood

Paul Lewis - Workshop on Molecular Evolution 2016

Stepping Stone
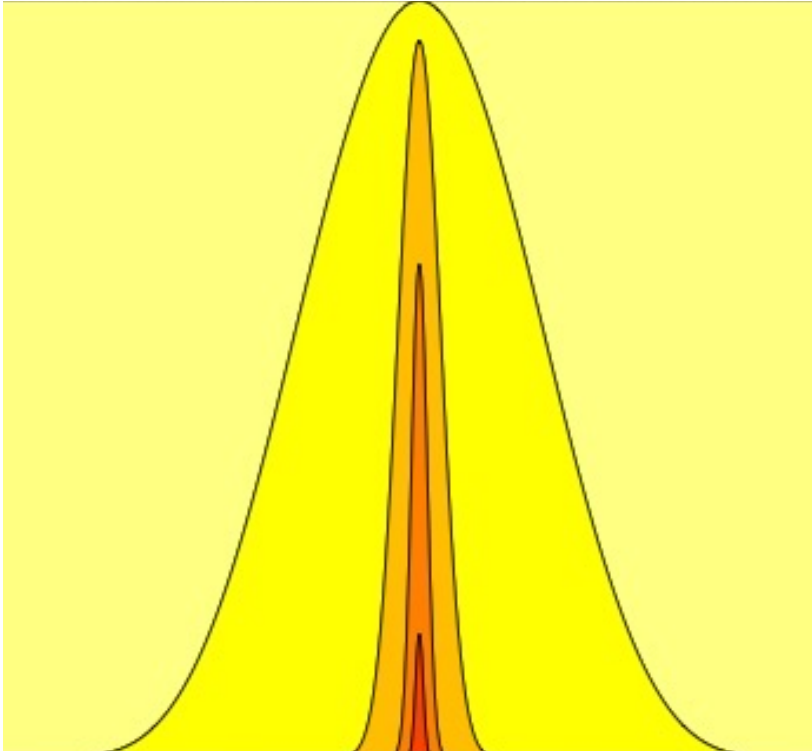
Keep estimating smaller and smaller sections until you get down to the marginal likelihood

Paul Lewis - Phyloseminar

Exercise 2

Check which of the two models you chose for exercise 1 better than the other using a model selection approach.
Can you already predict which one you think might be better?