

Model averaging, selection and adequacy

Taming the BEAST in London

Louis du Plessis

Department of Zoology, University of Oxford

July 2017



Introduction

Model selection

Model averaging

Model adequacy

References

What goes into a BEAST model?

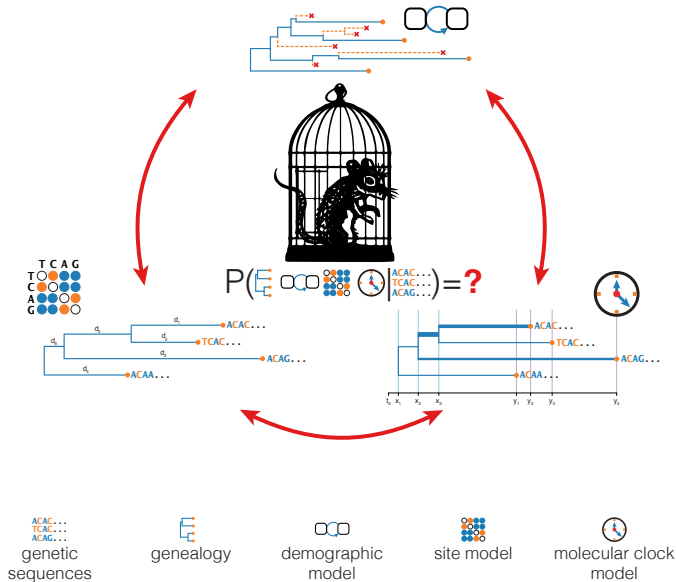
Introduction

Model selection

Model averaging

Model adequacy

References



Model consists of:

- ▶ Tree prior
- ▶ Site model
- ▶ Clock model
- ▶ Priors
- ▶ Hyperpriors

How can we decide which model components to use?

- ▶ **Model selection:** Which model is better?
- ▶ **Model adequacy:** Is the model any good?
- ▶ **Model averaging:** If you don't want to choose...

Posterior:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Introduction

Model selection

Model averaging

Model adequacy

References

Posterior:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Should actually write:

$$p(\theta|D, M) = \frac{p(D|M, \theta)p(\theta|M)}{p(D|M)}$$

- ▶ D: Data
- ▶ M: Model
- ▶ θ : Particular parameterization of model M

Posterior:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Should actually write:

$$p(\theta|D, M) = \frac{p(D|M, \theta)p(\theta|M)}{p(D|M)}$$

- ▶ D: Data
- ▶ M: Model
- ▶ θ : Particular parameterization of model M

$P(D|M)$ is the marginal probability of the data under model M

Need to calculate the marginal likelihood of the data to do Bayesian model selection:

$$P(D|M) = \int_{\theta} P(D|M, \theta) P(\theta|M) d\theta$$

Introduction

Model selection

Model averaging

Model adequacy

References

Need to calculate the marginal likelihood of the data to do Bayesian model selection:

$$P(D|M) = \int_{\theta} P(D|M, \theta) P(\theta|M) d\theta$$

Calculate Bayes factors:

$$BF_{M_1, M_2} = \frac{P(D|M_1)}{P(D|M_2)}$$

Marginal likelihood of the data

[Introduction](#)[Model selection](#)[Model averaging](#)[Model adequacy](#)[References](#)

Need to calculate the marginal likelihood of the data to do Bayesian model selection:

$$P(D|M) = \int_{\theta} P(D|M, \theta)P(\theta|M)d\theta$$

Calculate Bayes factors:

$$BF_{M_1, M_2} = \frac{P(D|M_1)}{P(D|M_2)}$$

- ▶ Models do **not** need to be nested!
- ▶ Cannot compare models on **different** datasets!
- ▶ $P(D|M)$ is **difficult** to calculate!

In Tracer:

- ▶ HME (Harmonic mean estimate)
- ▶ AIC_M

Introduction

Model selection

Model averaging

Model adequacy

References

In Tracer:

- ▶ HME (Harmonic mean estimate)
- ▶ AIC_M

“The total unsuitability of the harmonic mean estimator should have been apparent within an hour of its discovery.” - Radford Neal (<http://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever>)

In Tracer:

- ▶ HME (Harmonic mean estimate)
- ▶ AIC_M

“The total unsuitability of the harmonic mean estimator should have been apparent within an hour of its discovery.” - Radford Neal (<http://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever>)

Only use the AIC_M if you are short on time!

In Tracer:

- ▶ HME (Harmonic mean estimate)
- ▶ AIC_M

“The total unsuitability of the harmonic mean estimator should have been apparent within an hour of its discovery.” - Radford Neal (<http://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever>)

Only use the AIC_M if you are short on time!

In BEAST and BEAST2:

- ▶ Stepping stone
- ▶ Path-sampling

In Tracer:

- ▶ HME (Harmonic mean estimate)
- ▶ AIC_M

“The total unsuitability of the harmonic mean estimator should have been apparent within an hour of its discovery.” - Radford Neal (<http://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever>)

Only use the AIC_M if you are short on time!

In BEAST and BEAST2:

- ▶ Stepping stone
- ▶ Path-sampling

Stepping stone and path-sampling are the best that is currently available!

Model selection in Tracer

Model averaging

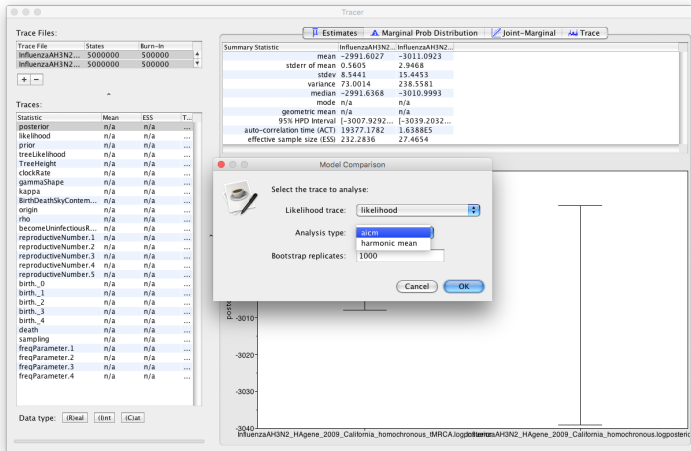
Introduction

Model selection

Model averaging

Model adequacy

References



Model selection in BEAST2

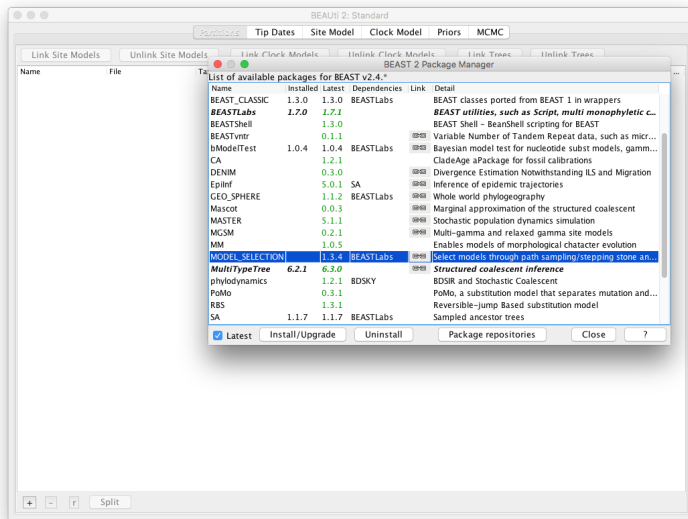
Introduction

Model selection

Model averaging

Model adequacy

References



Model selection in BEAST2

Model averaging

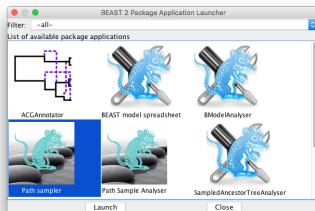
Introduction

Model selection

Model averaging

Model adequacy

References



Marginalization:

$$P(\phi) = \int_{\theta} P(\phi, \theta) d\theta = \int_{\theta} P(\phi|\theta)P(\theta)d\theta$$

- ▶ We can get rid of parameters we are not interested in by integrating them out!
- ▶ Effectively taking an average across all possible values of the parameter
- ▶ Estimates remaining parameters while taking into account uncertainty in nuisance parameters

Marginalization:

$$P(\phi) = \int_{\theta} P(\phi, \theta) d\theta = \int_{\theta} P(\phi|\theta)P(\theta)d\theta$$

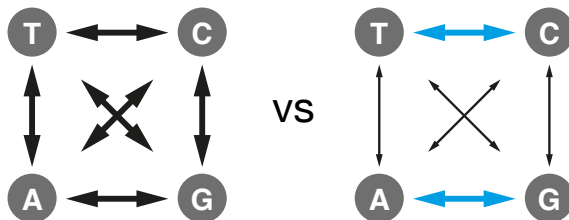
- ▶ We can get rid of parameters we are not interested in by integrating them out!
- ▶ Effectively taking an average across all possible values of the parameter
- ▶ Estimates remaining parameters while taking into account uncertainty in nuisance parameters

e.g. BEAST naturally takes into account phylogenetic uncertainty by integrating over all possible tree topologies!

We can actually integrate out components of the model in the same way! (nuisance models?)

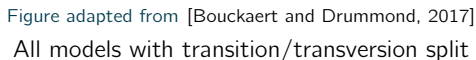
Reversible-jump MCMC (RJMCMC):

- ▶ Specify the state space of different models
- ▶ Need to tell your MCMC sampler how to step between different models
- ▶ Not easy to implement correctly (see [Green, 1995])



Stepping from JC69 to K80 means introducing an extra parameter

- Introduction
- Model selection
- Model averaging
- Model adequacy
- References



All time-
reversible
models

with/without
estimated
frequencies

with/without
Gamma rate
heterogeneity

with/without
invariant
sites

203

× 2

× 2

× 2

= 1,624 models!

All time-reversible models	with/without estimated frequencies	with/without Gamma rate heterogeneity	with/without invariant sites
203	× 2	× 2	× 2

= 1,624 models!

- ▶ BModelTest takes nucleotide model uncertainty into account
- ▶ The estimated rates are the average rates averaged over all models in the set of models allowed
- ▶ Support for a particular model is proportional to the time the chain spends in that model

(courtesy of David Duchêne)

Is the model actually any good?

Introduction

Model selection

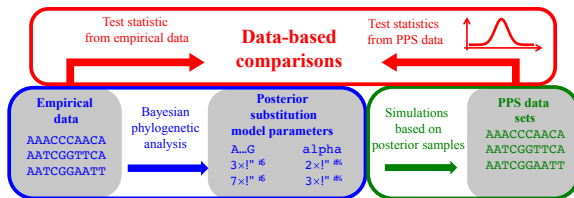
Model averaging

Model adequacy

References

(courtesy of David Duchêne)

Is the model actually any good? **Basic idea:**



Simulated data with estimated parameter should be similar to the empirical data

- Bouckaert, R. R. and Drummond, A. J. (2017). bmodeltest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1):42.
- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Introduction

Model selection

Model averaging

Model adequacy

References