(Bayesian) Model Testing

adapted from the Bodega Bay course on Phylogenetic Inference lecture by Brian R. Moore

SEBASTIAN HÖHNA

DEPARTMENT OF EVOLUTION & ECOLOGY, UC DAVIS
DEPARTMENT OF STATISTICS, UC BERKELEY

Model-Based Phylogenetic Inference

Model-based inference is based on the model

Model-Based Phylogenetic Inference

Model-based inference is based on the model

1. Model specification

   model selection

Model-Based Phylogenetic Inference

Model-based inference is based on the model

1. Model specification

    model selection

    model adequacy

Model-Based Phylogenetic Inference

Model-based inference is based on the model

1. Model specification

    model selection

    model adequacy

    model uncertainty/averaging

Model-Based Phylogenetic Inference

Model-based inference is based on the model

1. Model specification

   model selection

   model adequacy

   model uncertainty/averaging

2. Estimating under the model

   likelihood optimization

Model-Based Phylogenetic Inference

Model-based inference is based on the model

1. Model specification

    model selection

    model adequacy

    model uncertainty/averaging

2. Estimating under the model

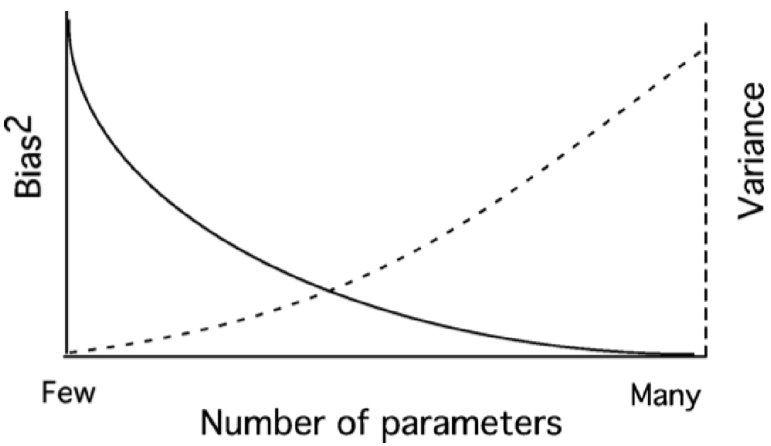
    likelihood optimization

    MCMC simulation

Model Specification Issues

Model selection, adequacy, and related issues

The model is central to parametric estimation of phylogeny: an under parameterized model will cause estimates to be biased (e.g., under estimation of branch lengths, topological error, inflated estimates of nodal support...); however, an over parameterized model will inflate estimation error (error variance, etc.).



Frequentist:

- Probability is *objective* and refers to the relative long term frequency.
- Parameters are all fixed and unknown constants.
- Produces point estimates of the parameters.
- Uncertainty is given by confidence intervals (CI).
- Maximization of hidden/latent parameters.

Bayesian:

- Probability is *subjective* and refers to the degree of belief.
- Parameters are are random variables.
- Produces posterior probability distribution (PP).
- Uncertainty is given by credible intervals (CI).
- Marginalization of hidden/latent parameters.

D = Data
θ = Model parameters

Posterior distribution

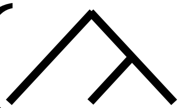
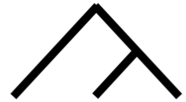
Prior distribution

"Likelihood"

$$f(\theta \mid D) = \frac{f(\theta)\,f(D \mid \theta)}{\int f(\theta)\,f(D \mid \theta)\,\mathrm{d}\theta}$$

Normalizing constant

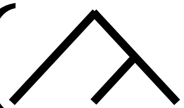
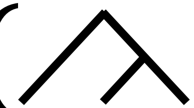
Likelihood function:

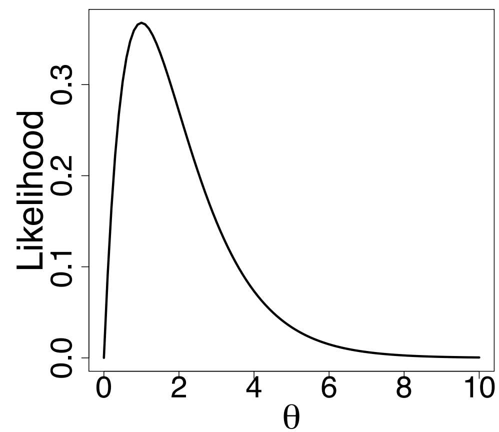
L( | ) = P( |  , $\pi,\gamma,\lambda,\mu,\ldots$)

Estimate: max L( | )

But what is the uncertainty in our estimate?

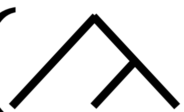
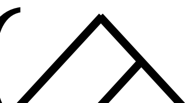
Profile Likelihood:

L( | ) = $\max_{\Theta}$(L( |  ,Θ))



Likelihood curve of Θ for a given tree

Marginal Likelihood:

L( | ) = $\int$ L( |  ,Θ) dΘ

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

1. Hierarchical Likelihood Ratio Test (hLRT)

Compare the ratio of maximum likelihood scores under a null (restricted) model and an alternative (more general) nested model

$$\Delta = 2(\ln L_1 - \ln L_0)$$

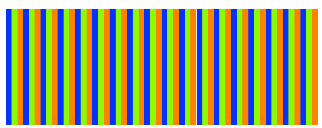
The statistic is (approximately) distributed as a Chi-square random variable with df equal to the difference in the number of free parameters in the two nested models.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

Original data matrix



Candidate models | ML estimation | ML estimates | $hLRT_{obs}$

M_0 ⇒ $\ln L_0$

M_1 ⇒ $\ln L_1$

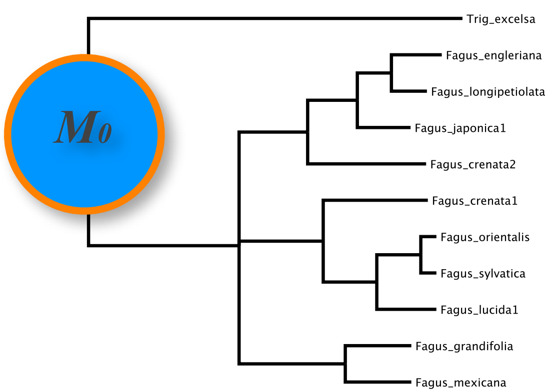
$$\Delta = 2(\ln L_1 - \ln L_0)$$

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

ML
estimates

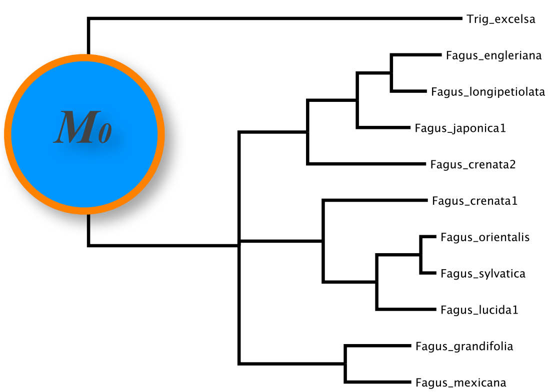


Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

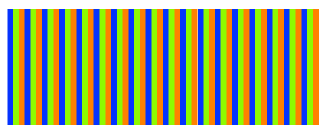
1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

ML estimates



Simulated data matrix

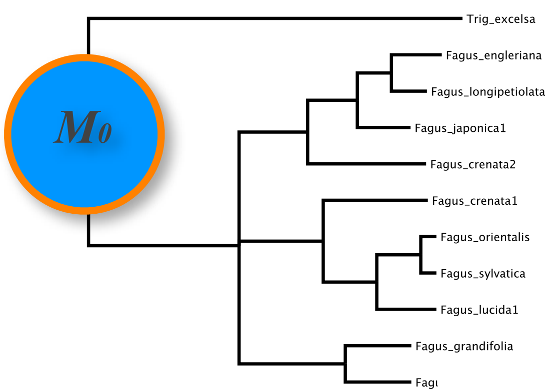


Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

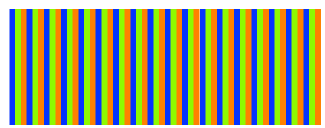
1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

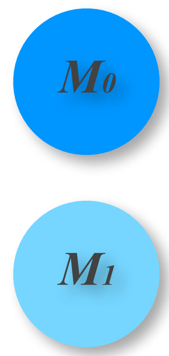
ML estimates



Simulated data matrix



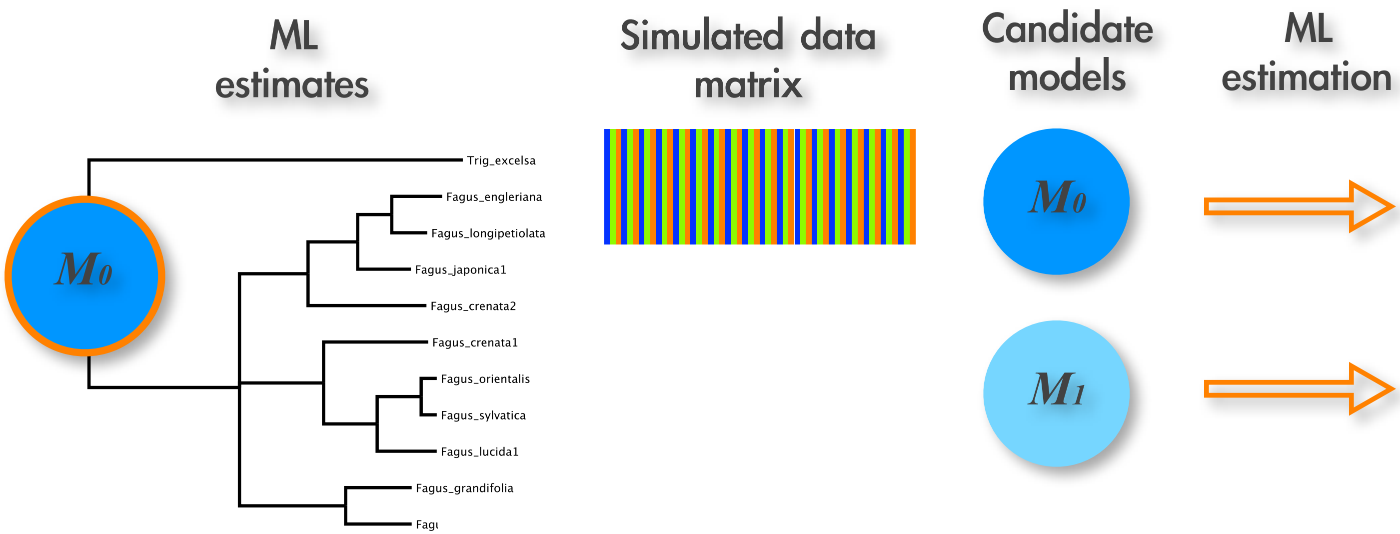
Candidate models



Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

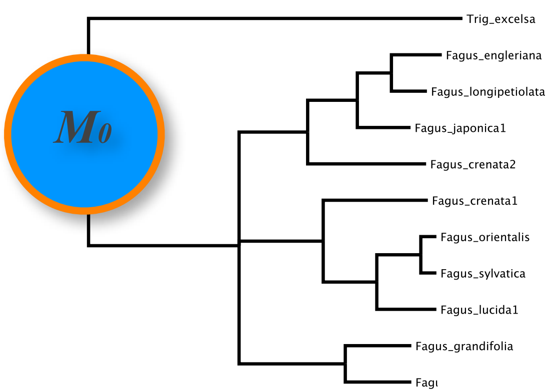


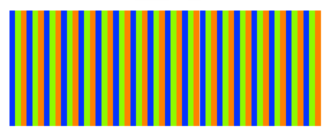
Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

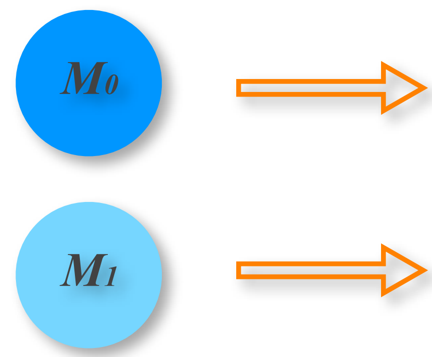
1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

ML estimates	Simulated data matrix	Candidate models	ML estimation	ML estimates	hLRT_{sim}







$\ln L_0$

$$\Delta = 2(\ln L_1 - \ln L_0)$$

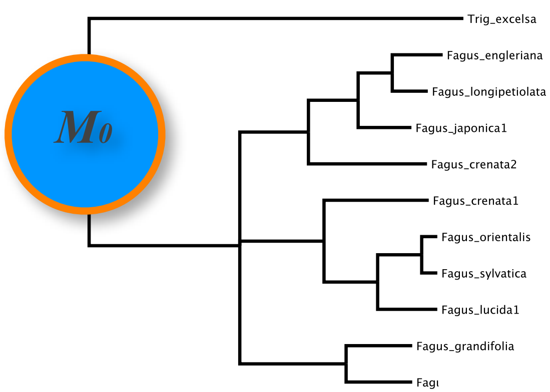
$\ln L_1$

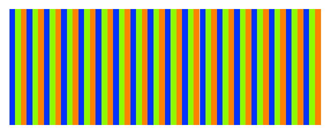
Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

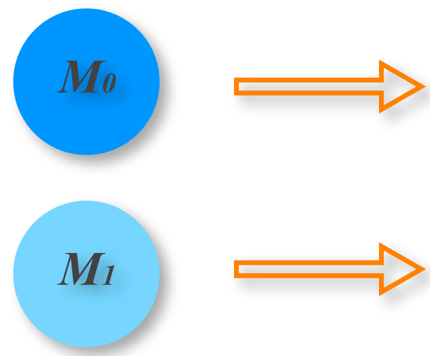
1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

ML estimates	Simulated data matrix	Candidate models	ML estimation	ML estimates	$hLRT_{sim}$







$\ln L_0$

$$\Delta = 2(\ln L_1 - \ln L_0)$$

$\ln L_1$

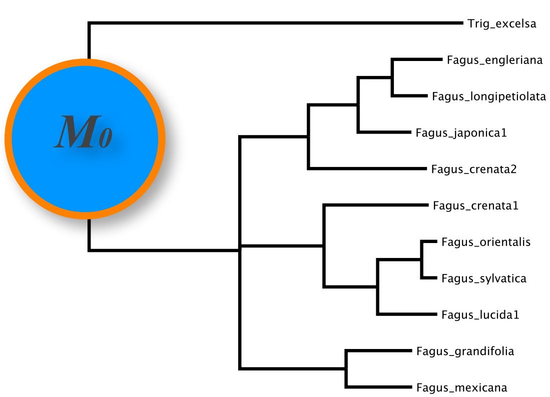
repeat many times...

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

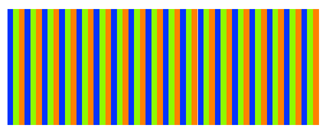
1. Hierarchical Likelihood Ratio Test (hLRT) for non-nested models

Parametric bootstrapping (Monte Carlo simulation)

ML estimates



Simulated data matrix

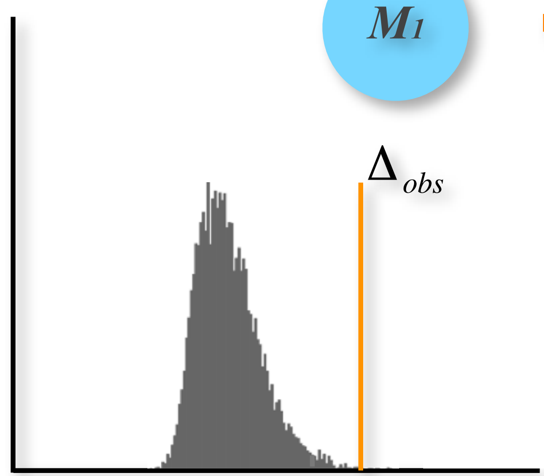


Candidate models | ML estimation | ML estimates | hLRT$_{sim}$

M_0 ⇒ $\ln L_0$

$$\Delta = 2(\ln L_1 - \ln L_0)$$

M_1 ⇒ $\ln L_1$



Null distribution of the test statistic

Reject null model
Fail to reject null model

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Score can be computed for a single model and compared to other candidate models:

$$\Delta AIC_i = AIC_i - \min AIC$$

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Score can be computed for a single model and compared to other candidate models:

$$\Delta AIC_i = AIC_i - \min AIC$$

Enables comparison of non-nested models.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Score can be computed for a single model and compared to other candidate models:

$$\Delta AIC_i = AIC_i - \min AIC$$

Enables comparison of non-nested models.

Avoids multiple test issues.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Score can be computed for a single model and compared to other candidate models:

$$\Delta AIC_i = AIC_i - \min AIC$$

Enables comparison of non-nested models.

Avoids multiple test issues.

Assumes ML estimates are known without error.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

2. Akaike Information Criterion (AIC)

Estimates the expected Kullback-Leibler information distance between a given model and the true, generating model (so smaller scores are better).

$$AIC_i = -2\ln L_i + 2p_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i).

Score can be computed for a single model and compared to other candidate models:

$$\Delta AIC_i = AIC_i - \min AIC$$

Enables comparison of non-nested models.

Avoids multiple test issues.

Assumes ML estimates are known without error.

Less biased toward more parameter-rich models than hLRT?

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\mathbf{ln}\,L_i + p_i\,\mathbf{ln}\,n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\ln L_i + p_i \ln n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\,\mathbf{ln}\,L_i + p_i\,\mathbf{ln}\,n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Enables comparison of non-nested models.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\ln L_i + p_i \ln n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Enables comparison of non-nested models.

Avoids multiple test issues.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\ln L_i + p_i \ln n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Enables comparison of non-nested models.

Avoids multiple test issues.

Assumes ML estimates are known without error.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\ln L_i + p_i \ln n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Enables comparison of non-nested models.

Avoids multiple test issues.

Assumes ML estimates are known without error.

Even less biased toward more parameter-rich models?

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

3. Bayesian Information Criterion (BIC)

A (crude) approximation of the marginal likelihood under the model, measuring the relative support for the model in the data (smaller values better).

$$BIC_i = -2\ln L_i + p_i \ln n_i$$

Attempts to balance model fit (the MLE under the estimation model) and error variance (the number of p parameters in model i multiplied by the n sample size in the data set).

Score can be computed for a single model and compared to other candidate models.

Enables comparison of non-nested models.

Avoids multiple test issues.

Assumes ML estimates are known without error.

Even less biased toward more parameter-rich models?

Assumes uniform prior over models and vague priors for parameters within models.

Maximum Likelihood Approaches for Model Selection/Hypothesis Testing

Relative stringency of ML-based model selection approaches

The approaches generally differ in their bias toward more parameter rich models:

$$hLRT > \Delta AIC > \Delta BIC$$

Different selection criteria may identify a different optimal model.

Bayesian Model testing

1) Model selection

2) Model adequacy testing

Bayesian Model testing

1) Model selection

Bayes factors

2) Model adequacy testing

Posterior Predictive Testing

Bayesian Model selection

$$BF = \frac{P(\mathrm{Data}|H_1)}{P(\mathrm{Data}|H_0)}$$

$$= \frac{\mathrm{Posterior}(H_1)}{\mathrm{Posterior}(H_0)} \times \frac{\mathrm{Prior}(H_0)}{\mathrm{Prior}(H_1)}$$

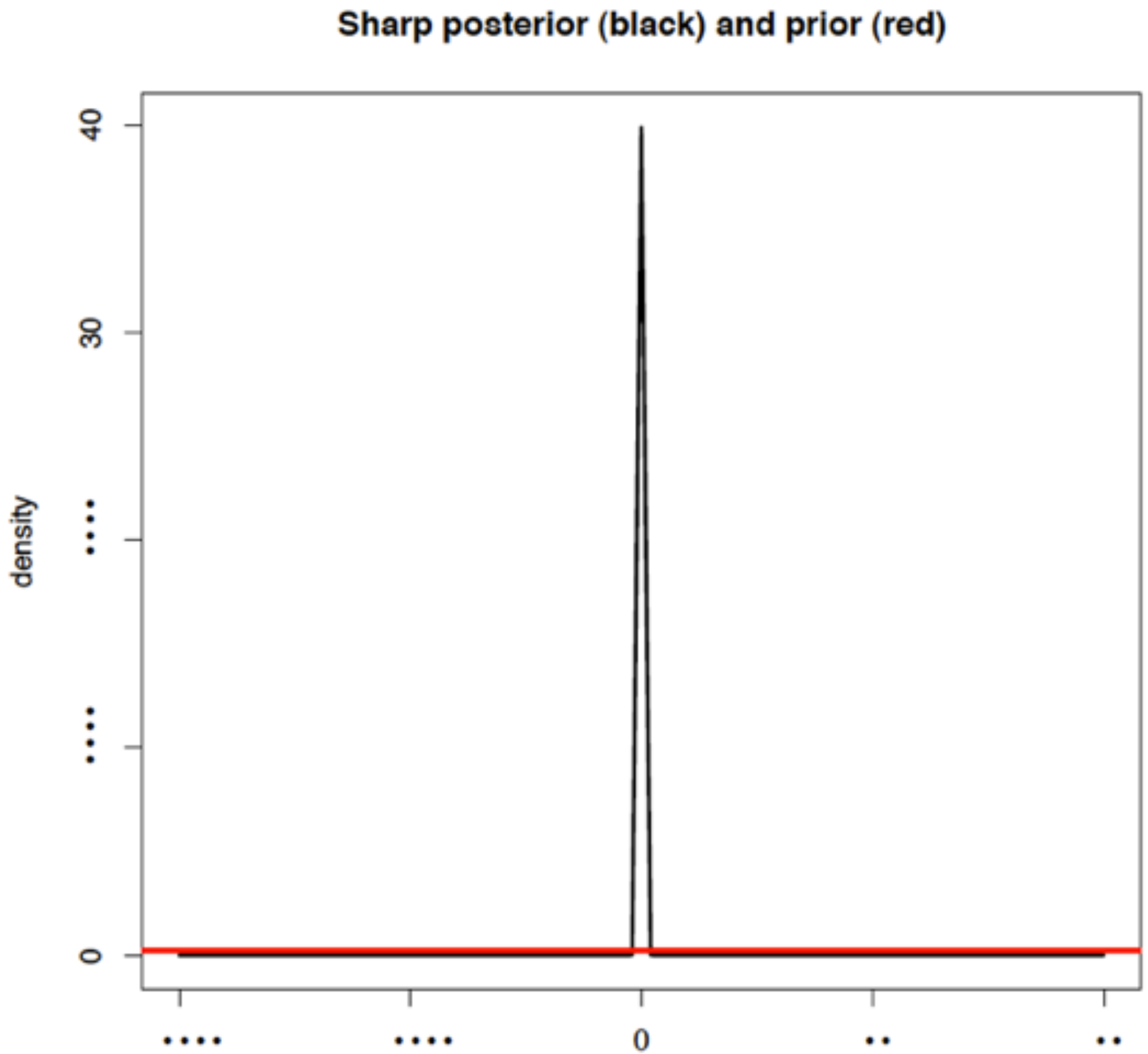
Interpreting Bayes factors

scale of evidence for Bayes factors	
Bayes factor	Interpretation
B.F. < 1/10	Strong evidence for Model 2
1/10 < B.F. < 1/3	Moderate evidence for Model 2
1/3 < B.F. < 1	Weak evidence for Model 2
1 < B.F. < 3	Weak evidence for Model 1
3 < B.F < 10	Moderate evidence for Model 1
B.F. > 10	Strong evidence for Model 1

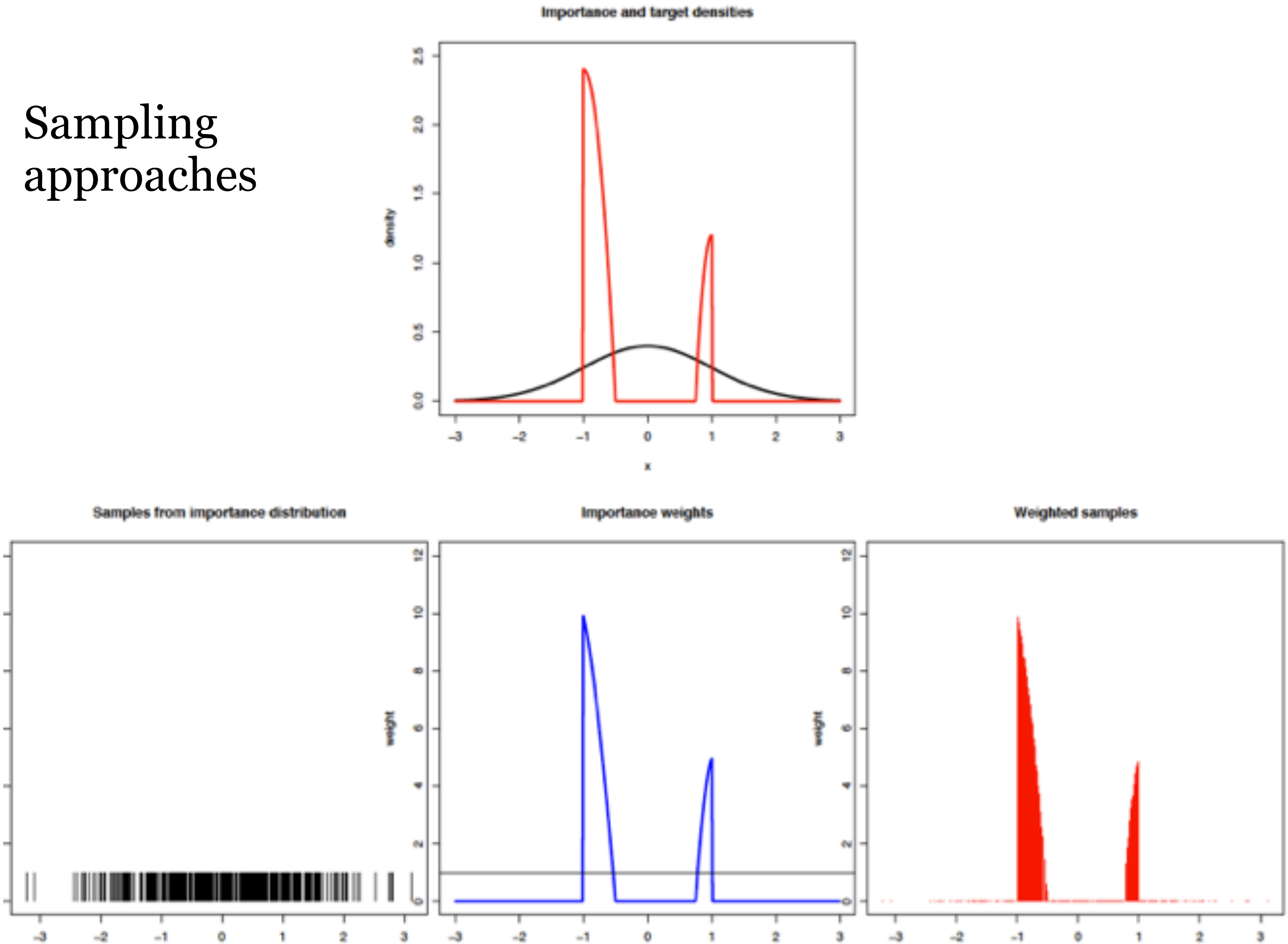
Marginal Likelihood

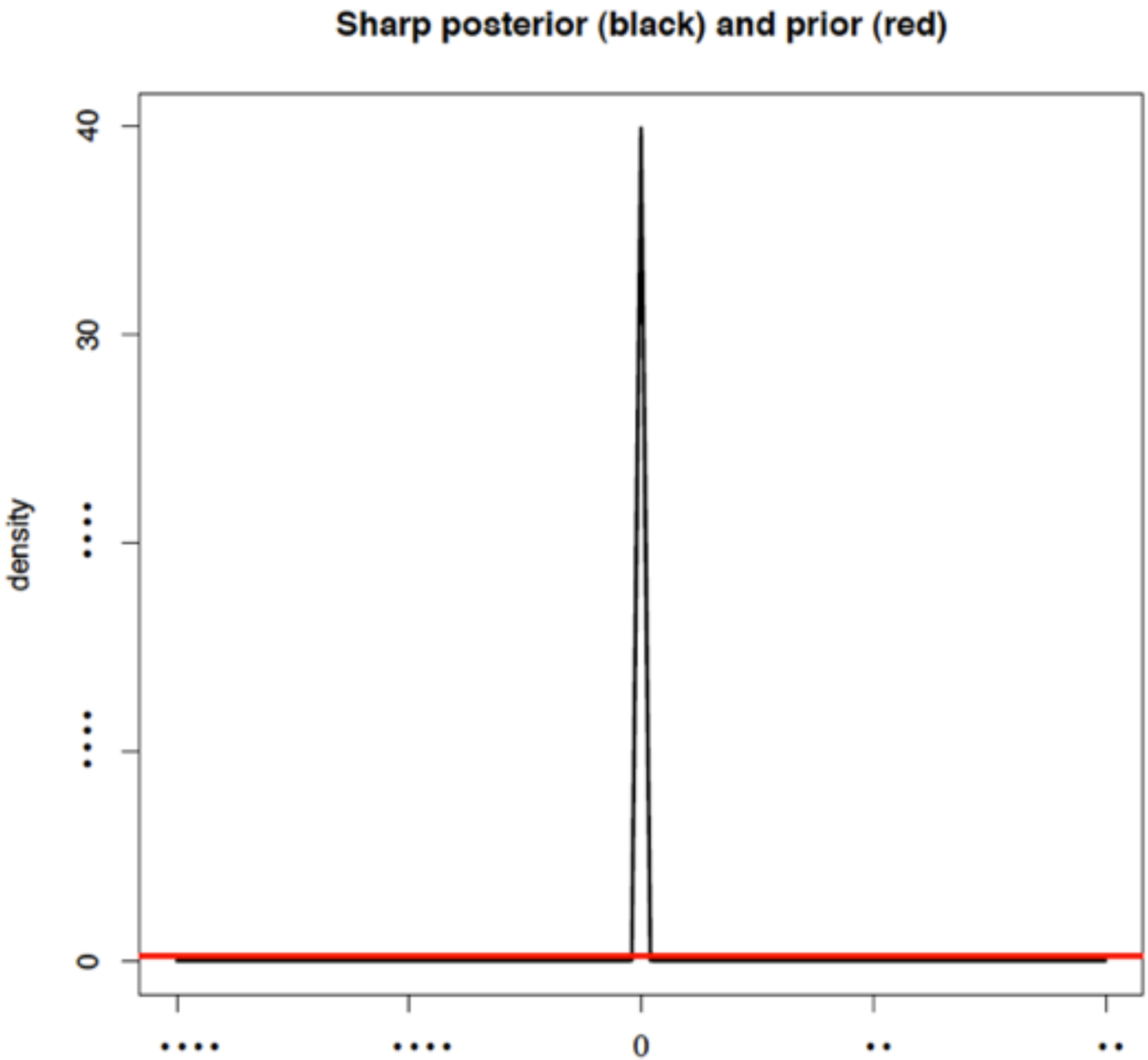
$$P(D|M) = \int P(D|\theta)p(\theta|M)d\theta$$

Probability of your data given the model, marginalized (integrate/summed) over all parameters.


Sharp posterior (black) and prior (red)
density
40
30
0
0

Sampling approaches




Sharp posterior (black) and prior (red)
density
40
30
0
0

Approximating Bayes factors

Harmonic mean estimator:

$$H = \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{-1}\right)^{-1} = \frac{1}{\frac{1}{n} \cdot \left(\frac{1}{x1} + \frac{1}{x2} + \cdots + \frac{1}{x_n}\right)} = \frac{n}{\frac{1}{x1} + \frac{1}{x2} + \cdots + \frac{1}{x_n}}.$$

x_i is the likelihood for sample i

Approximating Bayes factors

Reversible-Jump MCMC

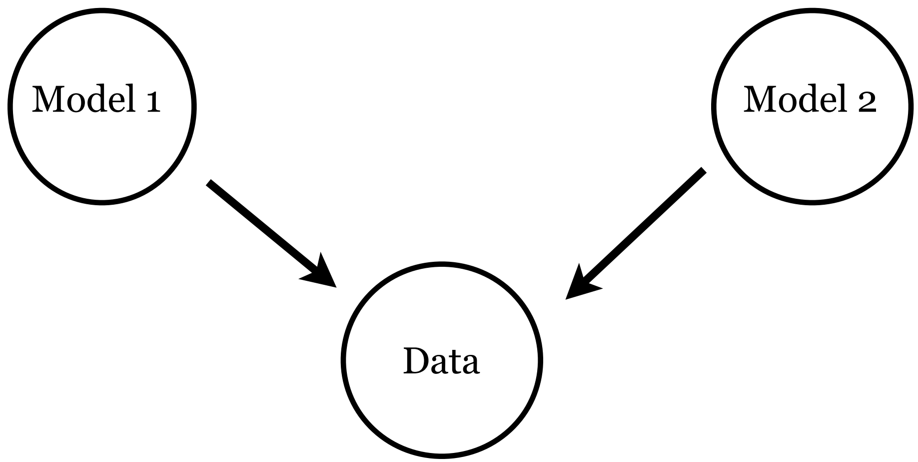
Model 1 ⟷ Model 2

Propose new model
with some values.
Accept/Reject using MCMC

Problem: Finding moves for proposing models is hard and model specific

Approximating Bayes factors

Mixture Models:



P(Data|M1,M2) = 0.5*P(Data|M1) + 0.5*P(Data|M2)

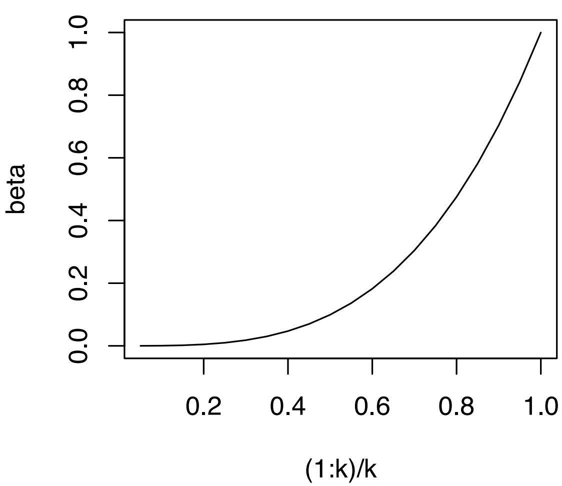
Use MCMC for accept/reject.

Sample M1 with p = P(Data|M1) / (P(Data|M1) + P(Data|M2))

Approximating Bayes factors

Stepping-Stone-Sampling:

Run an MCMC with the likelihood to the power of beta.

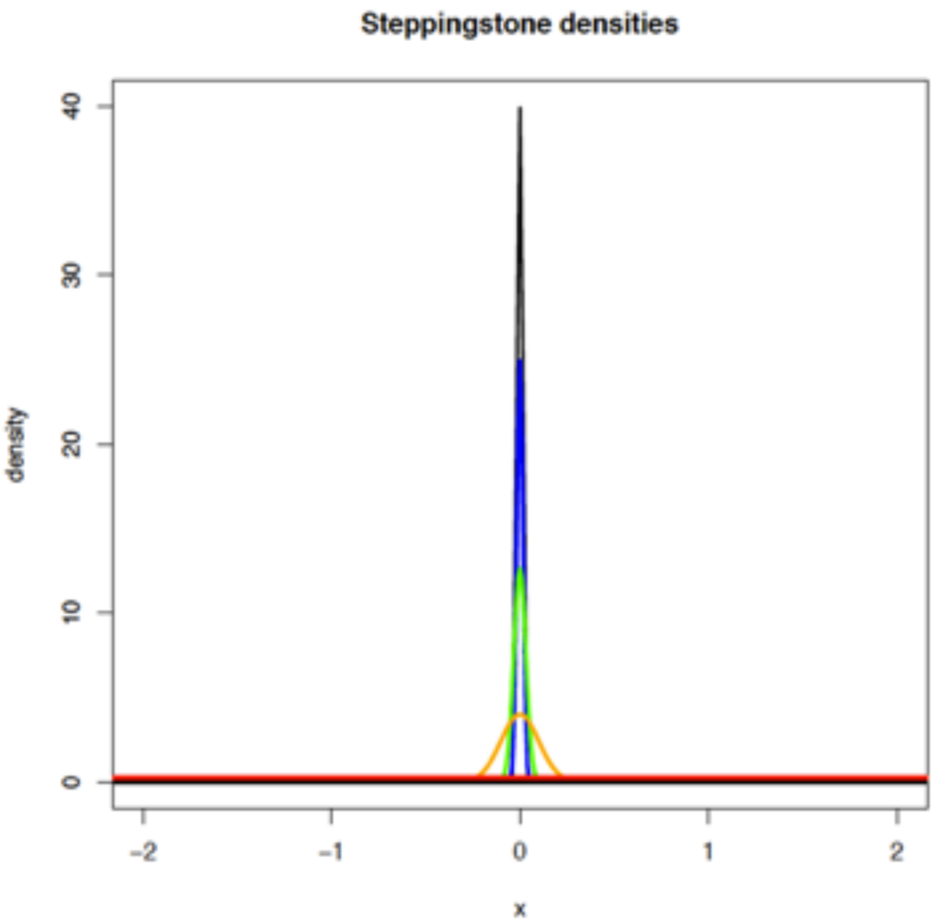


P(Θ|D) = P(D|Θ)^(beta) * P(Θ)

$$\mathbb{P}(D \mid M) = \left(\frac{\mathbb{P}(D|M)}{c_{0.38}}\right)\left(\frac{c_{0.38}}{c_{0.1}}\right)\left(\frac{c_{0.1}}{c_{0.01}}\right)\left(\frac{c_{0.01}}{1}\right)$$

Approximating Bayes factors

Path-Sampling (Stepping-Stone-Sampling):



Path-Sampling

- Path-sampling (PS) and Stepping-Stone-Sampling (SSS) use the same power posterior.
- Once you have run the power-posterior mcmc you can estimate both (PS and SSS)
- SSS is slightly more robust.
- Both are time-consuming, but converge towards the true marginal likelihood.

Summary

- Models are selected based on their marginal likelihood.

- The harmonic mean estimator is fast, but unreliable!

- The priors are VERY important for the marginal likelihood! Always check for the influence of the priors.

- Try estimating the marginal likelihood with a prior mean divided by 2 and multiplied by 2.