

A Brief Introduction to Diagnosing MCMC Performance

Brian R. Moore

Department of Evolution & Ecology
UC, Davis

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

MCMC Approximation of the Joint Posterior Probability Density

MCMC in theory and practice

MCMC in theory...

an appropriately constructed and adequately run chain is guaranteed to provide an arbitrarily precise description of the joint stationary density

MCMC in practice...

although a given sampler may work well in most cases, all samplers will fail in some cases, and is not guaranteed to work for any given case

Q. When do we know that the MCMC provides an accurate approximation for a given empirical analysis?

A.

NEVER!

MCMC Approximation of the Joint Posterior Probability Density

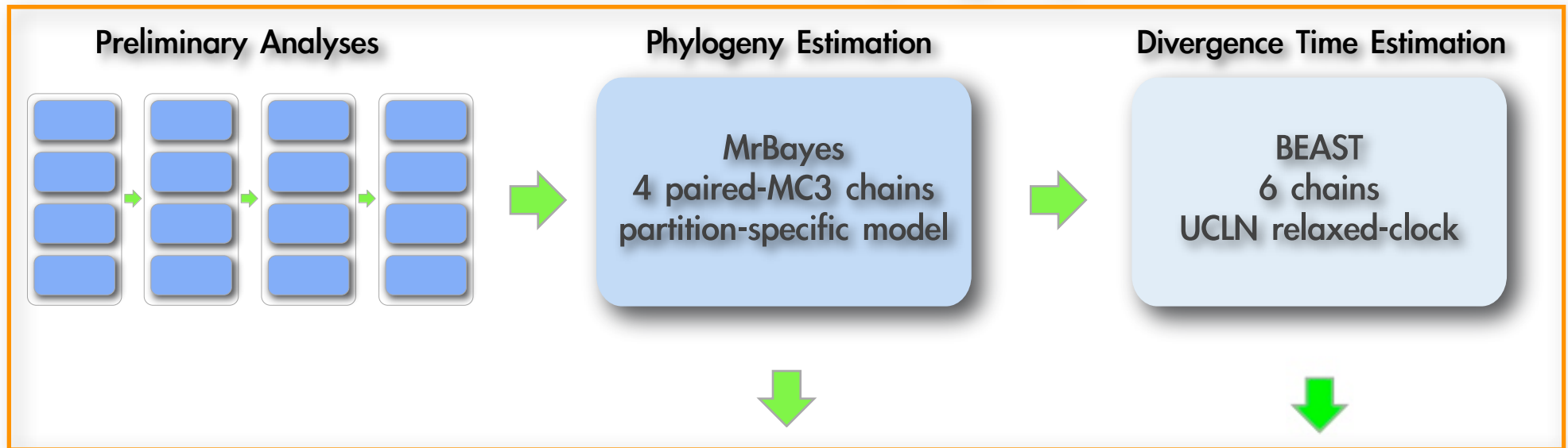
MCMC performance and OCD

It is not sufficient to merely be deeply concerned about MCMC performance...you need to be **completely obsessed** about it!
for **any** Bayesian inference based on MCMC
particularly for complex models/inference problems

Outline

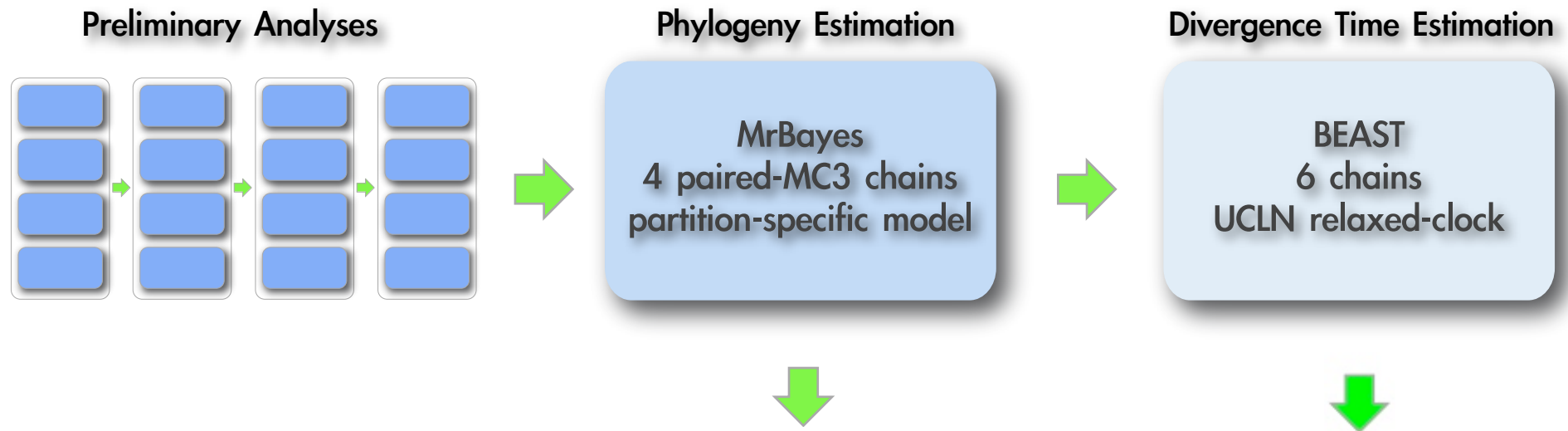
- 
- I. A review of where we've been and why
 - II. Diagnosing MCMC performance
 - III. Diagnostics based on single chains
 - IV. Diagnostics based on the prior
 - V. Diagnostics based on multiple chains

Statistical Estimation of Phylogeny: An Outline



	TOPOLOGY	CHRONOGRAM
Discrete Trait Evolution	BayesTraits	BayesTraits
Continuous Trait Evolution	misc. R	misc. R
Rates of Trait Evolution	misc R	misc R
Rates of Lineage Diversification	symmeTREE	misc. R

Statistical Estimation of Phylogeny: An Outline



	TOPOLOGY	CHRONOGRAM
Discrete Trait Evolution	BayesTraits	BayesTraits
Continuous Trait Evolution	misc. R	misc. R
Rates of Trait Evolution	misc R	misc R
Rates of Lineage Diversification	symmeTREE	misc. R

Bayesian Inference of Phylogeny in One Slide

$$\text{posterior probability} \quad \Pr(\text{Parameter} \mid \text{Data}) = \frac{\text{likelihood function} \quad \Pr(\text{Data} \mid \text{Parameter}) \times \text{prior probability} \quad \Pr(\text{Parameter})}{\text{marginal likelihood} \quad \Pr(\text{Data})}$$

I. Data

Assume an alignment, X , of N sites for S species: $X = (x_1, x_2, \dots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = (\tau_1, \tau_2, \dots, \tau_{(2S-5)!!})$
 branch lengths $\nu = (\nu_1, \nu_2, \dots, \nu_{(2S-3)})$

\sim Uniform

\sim Dirichlet (1, ..., 1)

2. Model of character change $\Phi = (\theta, \pi, \alpha, T)$

substitution rates $\theta = (\theta_{AC}, \theta_{AG}, \theta_{AT}, \theta_{CG}, \theta_{CT}, \theta_{GT}) \sim$ Dirichlet (1, 1, 1, 1, 1, 1)

stationary frequencies $\pi = (\pi_A, \pi_C, \pi_G, \pi_T) \sim$ Dirichlet (1, 1, 1, 1)

among-site rate variation $\alpha \sim$ Uniform (0.1, 100)

IV. Priors on parameters

III. Phylogenetic likelihood function

$$L(\tau, \nu, \Phi) = f(X \mid \tau, \nu, \Phi) = \sum_{j=1}^{(2S-5)!!} \int \prod_{k=1}^{2S-3} \prod_{i=1}^N f(x_i \mid \tau_j, \nu_k, \Phi) f(\nu_k, \Phi) d\nu_k d\Phi$$

V. Posterior Probability

$$f(\tau, \nu, \Phi \mid X) = \frac{f(X \mid \tau, \nu, \Phi) f(\tau, \nu, \Phi)}{f(X)}$$

Approximating the Joint Posterior Probability Density with MCMC

Samples from the chain approximate the joint posterior

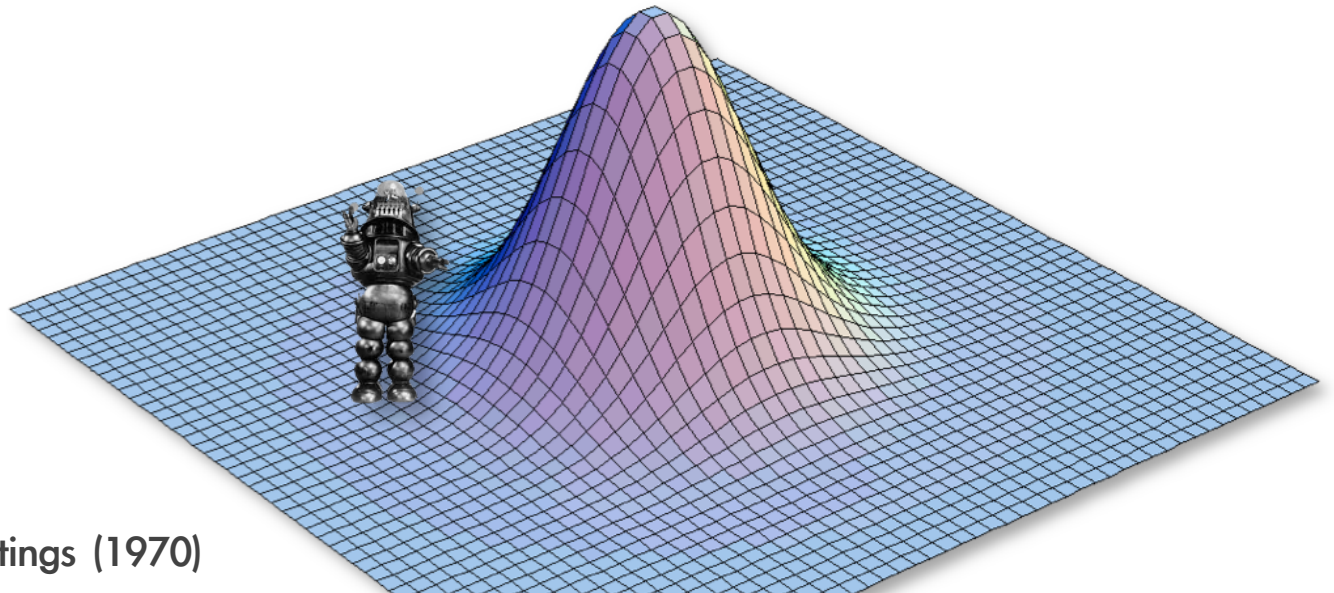
The frequency of sampled parameter values provides a valid estimate of the posterior probability of that parameter

We can query the joint posterior with respect to any individual parameter of interest: the marginal posterior probability

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:

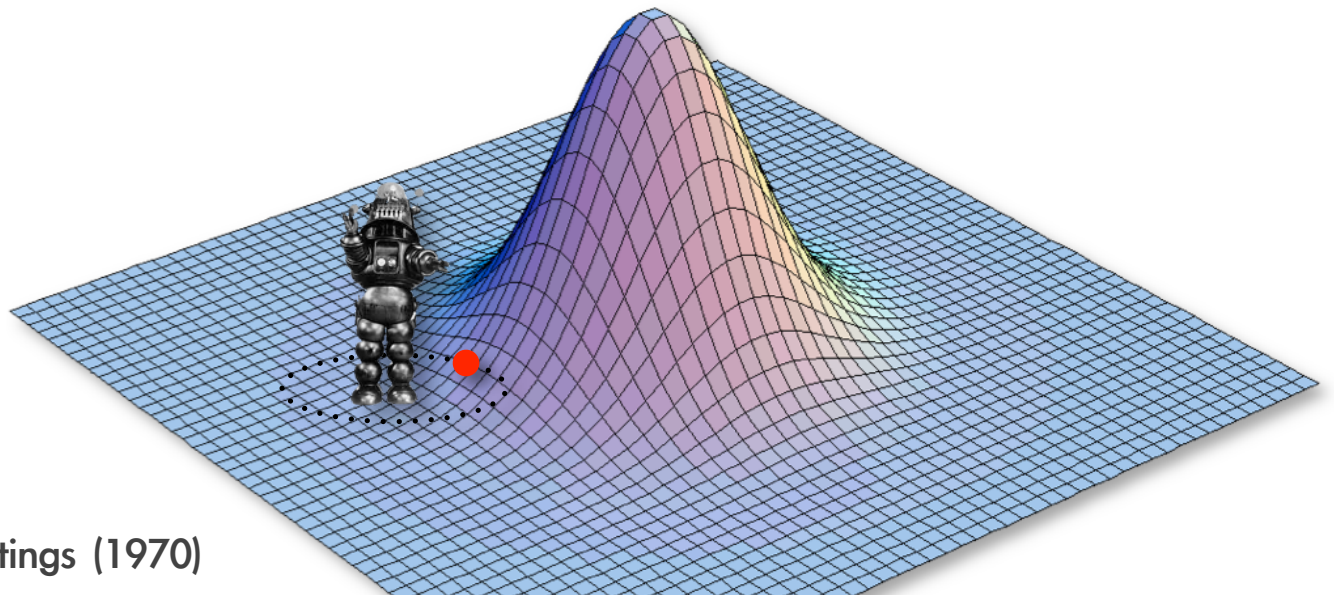


Metropolis et al. (1953); Hastings (1970)

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:



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

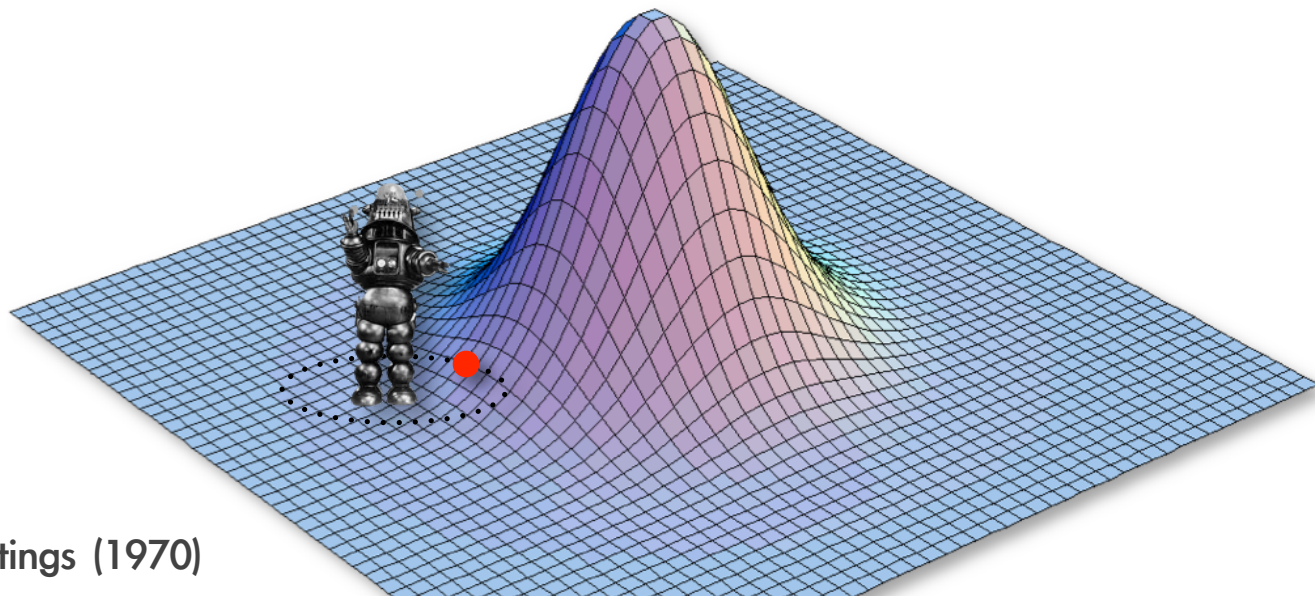
Approximating the Joint Posterior Probability Density with MCMC

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

$$\Pr[\text{Accept}] = 1$$

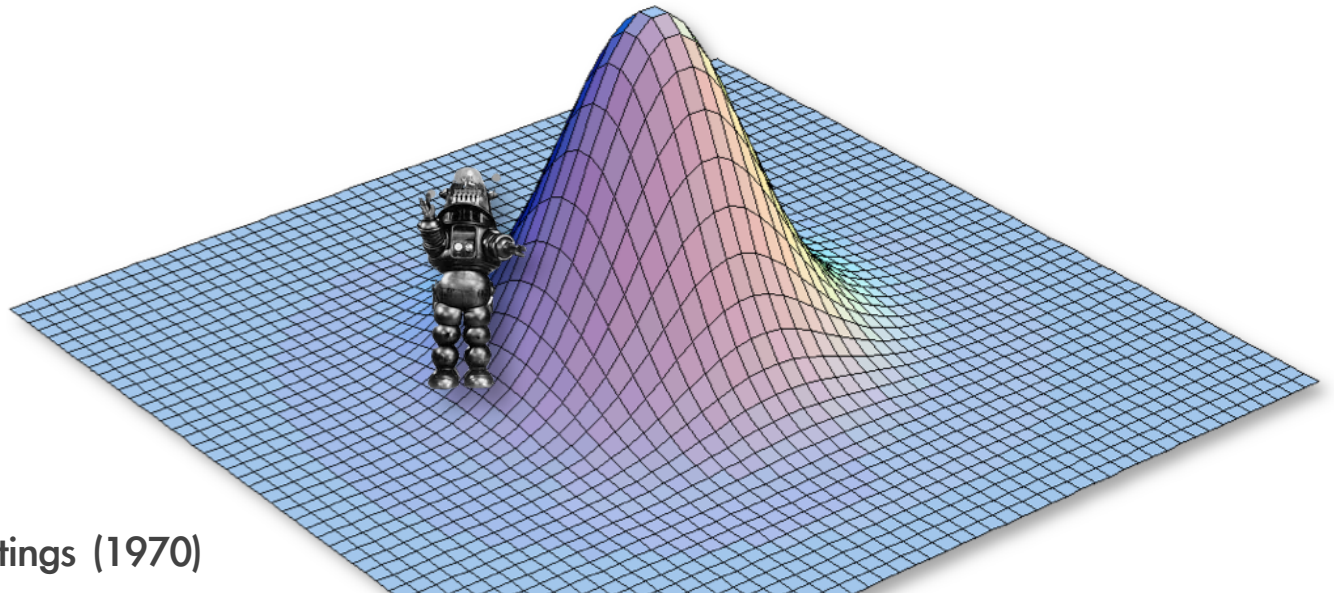


Approximating the Joint Posterior Probability Density with MCMC

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



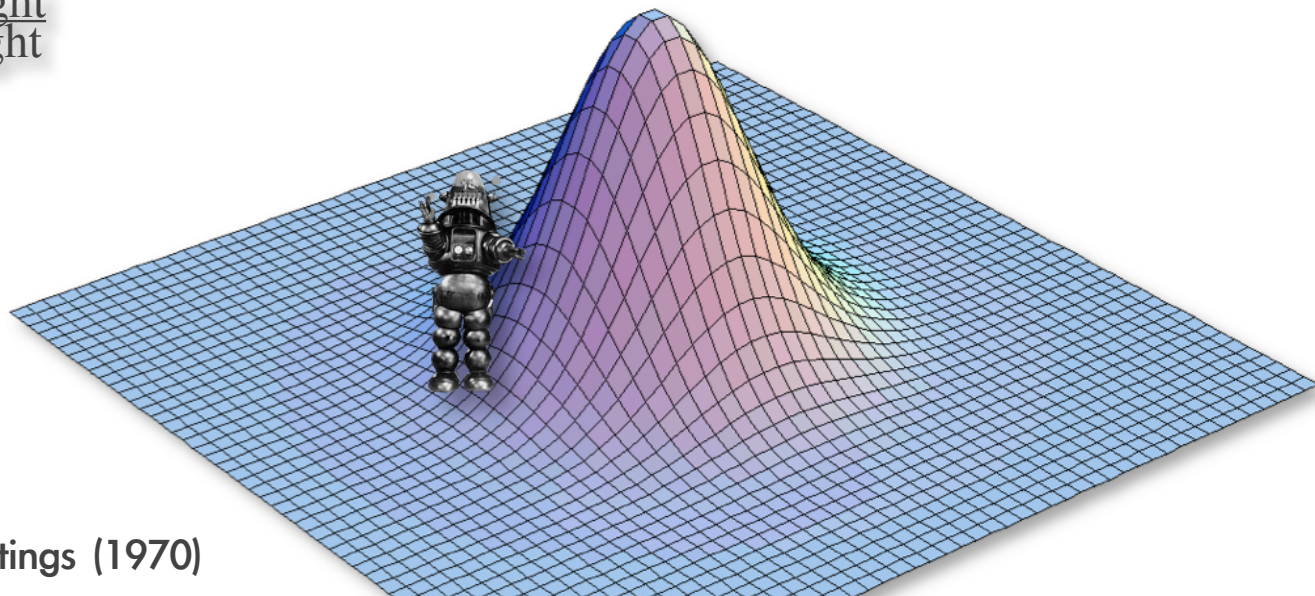
Approximating the Joint Posterior Probability Density with MCMC

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, $U[0,1]$

$$\text{Pr}[\text{Accept}] = \frac{\text{new height}}{\text{old height}}$$

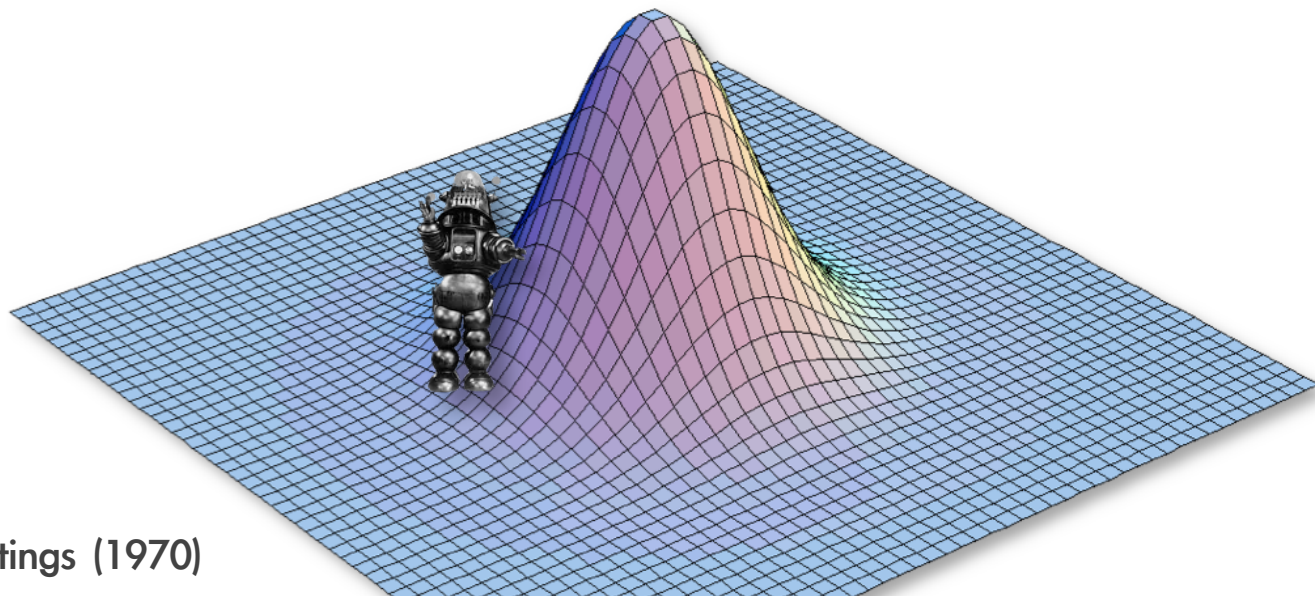


Approximating the Joint Posterior Probability Density with MCMC

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 $\Pr[A \rightarrow B] = \Pr[B \rightarrow A]$

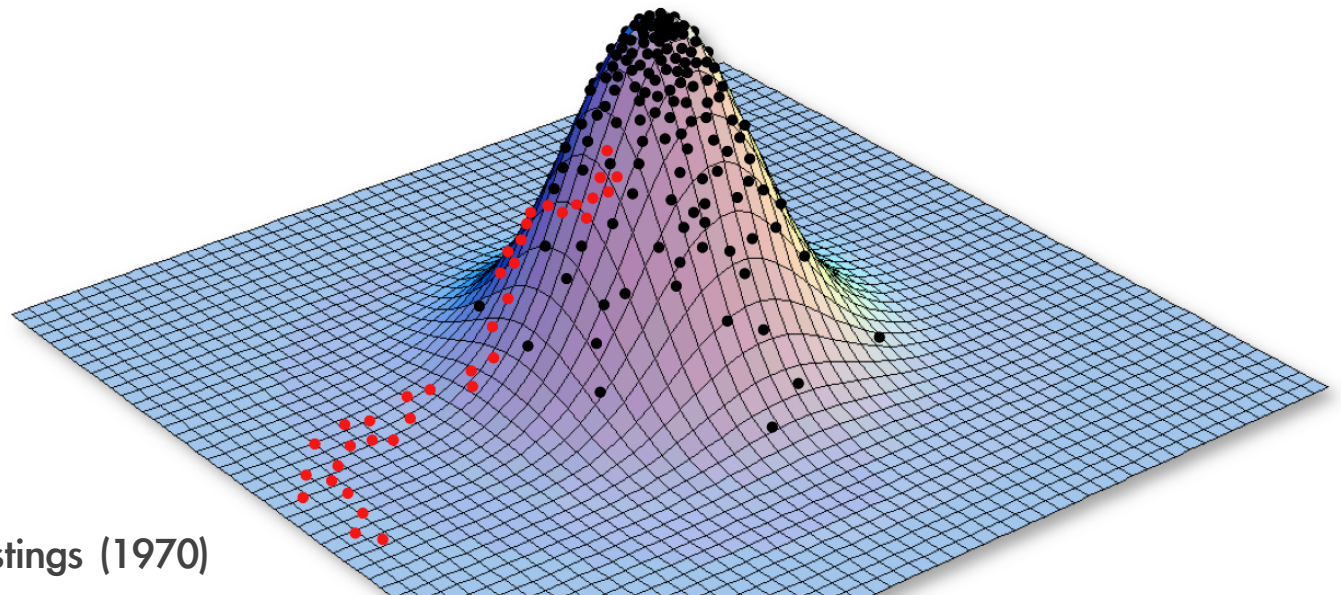


Approximating the Joint Posterior Probability Density with MCMC

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 $\Pr[A \rightarrow B] = \Pr[B \rightarrow A]$



Approximating the Joint Posterior Probability Density with MCMC

A slightly more formal description of the Metropolis-Hastings algorithm...

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = (\tau, \nu)$
2. Select a parameter to change according to its proposal probability, e.g.:

The MCMC sampler will use the following moves:

With prob.	Chain will change
2.86 %	param. 1 (tratio) with Dirichlet proposal
2.86 %	param. 2 (revmat) with Dirichlet proposal
2.86 %	param. 3 (revmat) with Dirichlet proposal
2.86 %	param. 4 (state frequencies) with Dirichlet proposal
2.86 %	param. 5 (state frequencies) with Dirichlet proposal
2.86 %	param. 6 (state frequencies) with Dirichlet proposal
2.86 %	param. 7 (state frequencies) with Dirichlet proposal
2.86 %	param. 8 (state frequencies) with Dirichlet proposal
2.86 %	param. 9 (gamma shape) with multiplier
2.86 %	param. 10 (gamma shape) with multiplier
2.86 %	param. 11 (gamma shape) with multiplier
2.86 %	param. 12 (gamma shape) with multiplier
2.86 %	param. 13 (gamma shape) with multiplier
2.86 %	param. 14 (prop. invar. sites) with beta proposal
2.86 %	param. 15 (rate multiplier) with Dirichlet proposal
14.29 %	param. 16 (topology and branch lengths) with LOCAL
42.86 %	param. 16 (topology and branch lengths) with extending TBR

Approximating the Joint Posterior Probability Density with MCMC

A slightly more formal description of the Metropolis-Hastings algorithm...

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = (\tau, \nu)$
2. Select a parameter to change according to its proposal probability
3. Propose a change to the selected parameter using the parameter-specific proposal mechanism that is:
 - (i) stochastic
 - (ii) irreducible
 - (iii) aperiodic

Approximating the Joint Posterior Probability Density with MCMC

A slightly more formal description of the Metropolis-Hastings algorithm...

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = (\tau, \nu)$
2. Select a parameter to change according to its proposal probability
3. Propose a change to the selected parameter using the parameter-specific proposal mechanism
4. Calculate the probability of accepting the proposed change:

$$R = \min \left[1, \underbrace{\frac{f(X | \Theta')}{f(X | \Theta)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{f(\Theta')}{f(\Theta)}}_{\text{prior ratio}} \cdot \underbrace{\frac{f(\Theta | \Theta')}{f(\Theta' | \Theta)}}_{\text{proposal ratio}} \right]$$

Approximating the Joint Posterior Probability Density with MCMC

A slightly more formal description of the Metropolis-Hastings algorithm...

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = (\tau, \nu)$
2. Select a parameter to change according to its proposal probability
3. Propose a change to the selected parameter using the parameter-specific proposal mechanism
4. Calculate the probability of accepting the proposed change
5. Generate a uniform random variable, $U[0,1]$, accept if $R > U$

Approximating the Joint Posterior Probability Density with MCMC

A slightly more formal description of the Metropolis-Hastings algorithm...

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = (\tau, \nu)$
2. Select a parameter to change according to its proposal probability
3. Propose a change to the selected parameter using the parameter-specific proposal mechanism
4. Calculate the probability of accepting the proposed change
5. Generate a uniform random variable, $U[0,1]$, accept if $R > U$
6. Repeat steps 2 - 5 an 'adequate' number of times

Outline

I. A review of where we've been and why

 II. Diagnosing MCMC performance

III. Diagnostics based on single chains

IV. Diagnostics based on the prior

V. Diagnostics based on multiple chains

Assessing MCMC Performance: Three Main Issues

1. Convergence

Has the chain (robot) successfully targeted the stationary distribution?

2. Mixing

Is the chain (robot) successfully integrating over the joint posterior probability?

3. Sampling intensity

Has the robot collected enough samples to adequately describe the posterior probability distribution?

Outline

- I. A review of where we've been and why
- II. Diagnosing MCMC performance
-  III. Diagnostics based on single chains
- IV. Diagnostics based on the prior
- V. Diagnostics based on multiple chains

Assessing MCMC Performance: Diagnostics Based on Single Runs

1. Convergence diagnostics

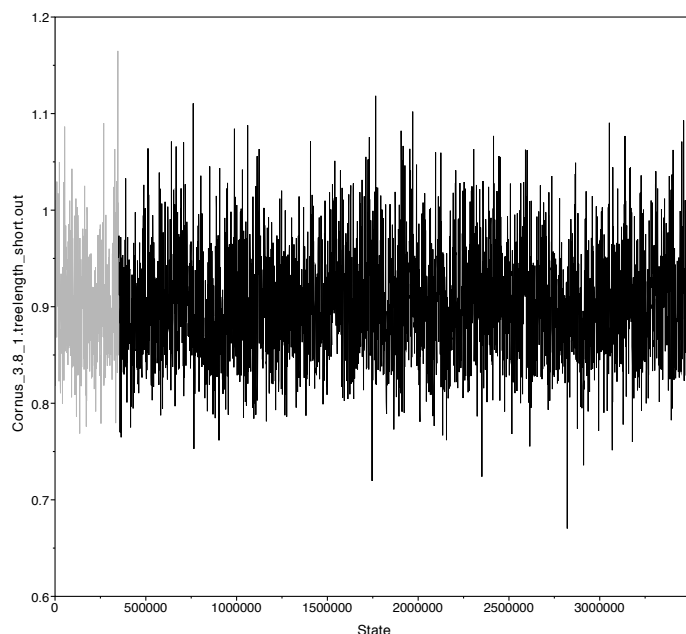
(i) Time-series plots of parameter estimates

- continuous parameters (e.g., substitution rates)--Tracer
 - some parameters are more reliable than others
 - steps may occur!

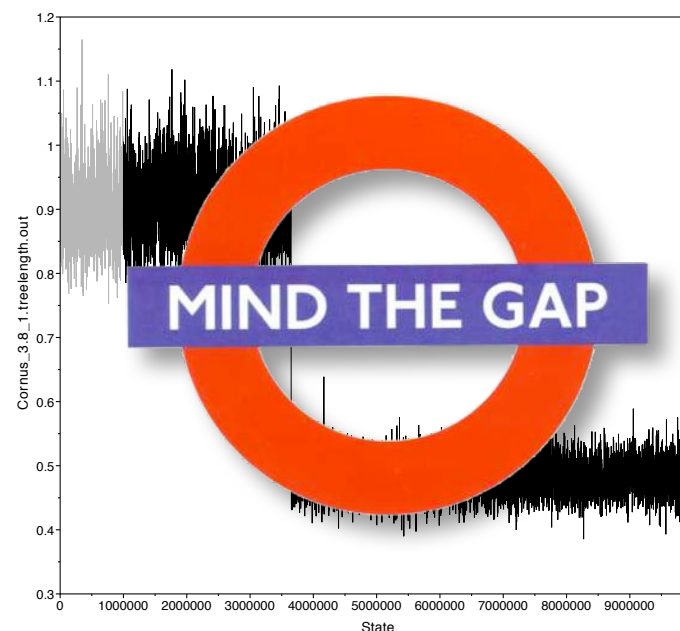
Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Tracer plots of tree-length at two stages of a single MrBayes run

bad convergence



better convergence



fast*

slow*

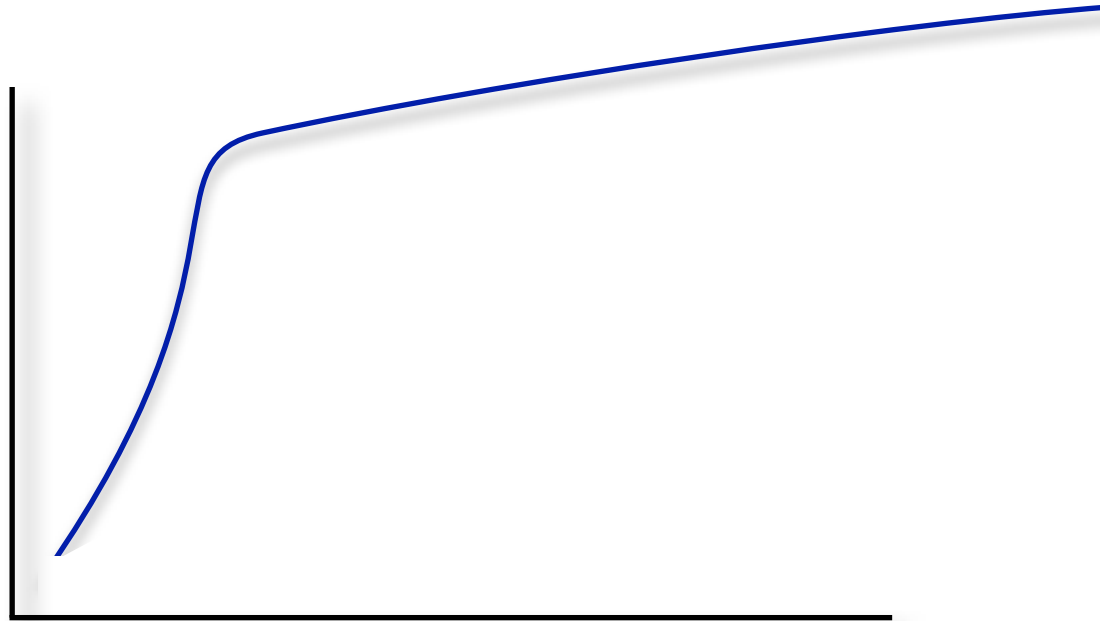


*somewhat data-set dependent

Assessing MCMC Performance: Diagnostics Based on Single Runs

MCMC pathologies

Slow convergence of time-series plots of parameter estimates

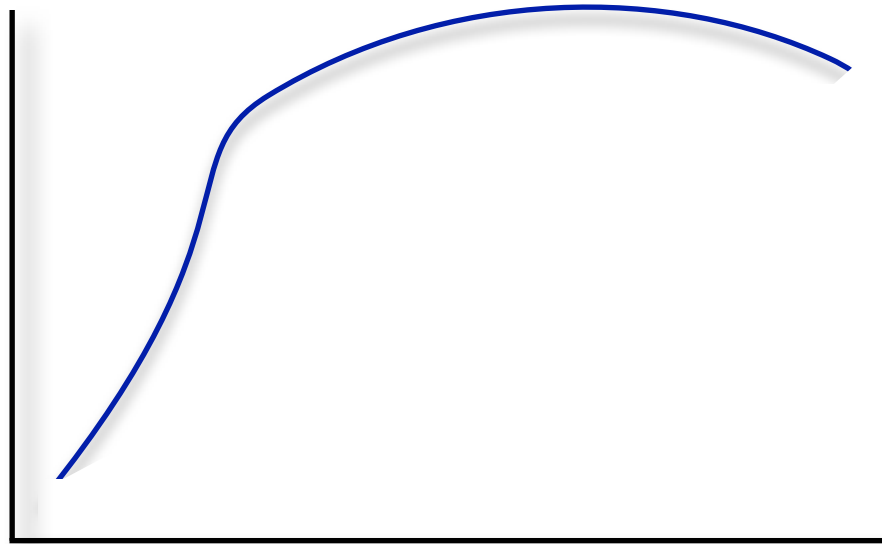


- under-parameterized model
- scale of tuning parameters too small (acceptance rates too high)
- inappropriate priors (e.g., tree length)

Assessing MCMC Performance: Diagnostics Based on Single Runs

MCMC pathologies

Decrease in time-series plot of $\ln L$ estimates ('burnout')



- over-parameterized model
- poorly chosen/unrealistic priors
- one or more weak parameters (marginal posteriors & posteriors similar)

Assessing MCMC Performance: Diagnostics Based on Single Runs

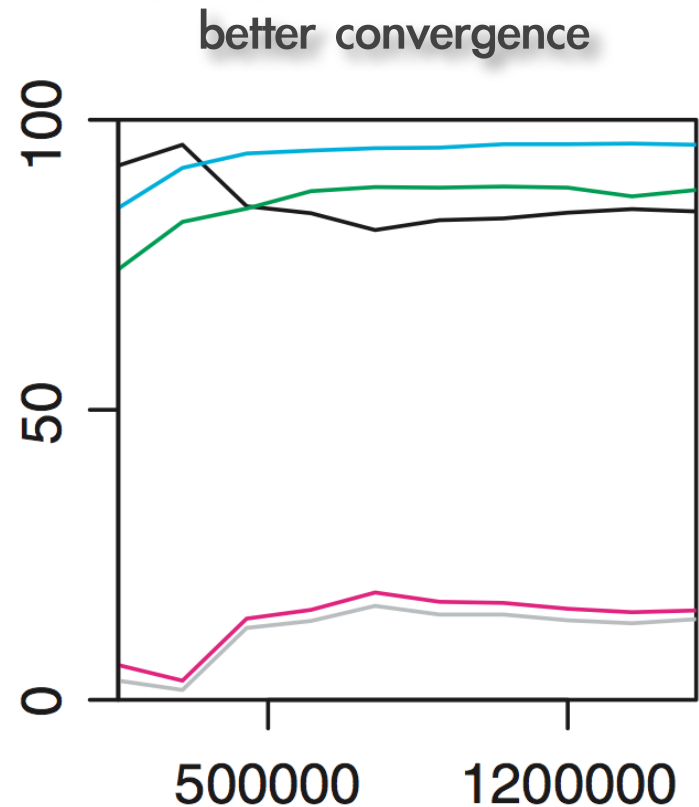
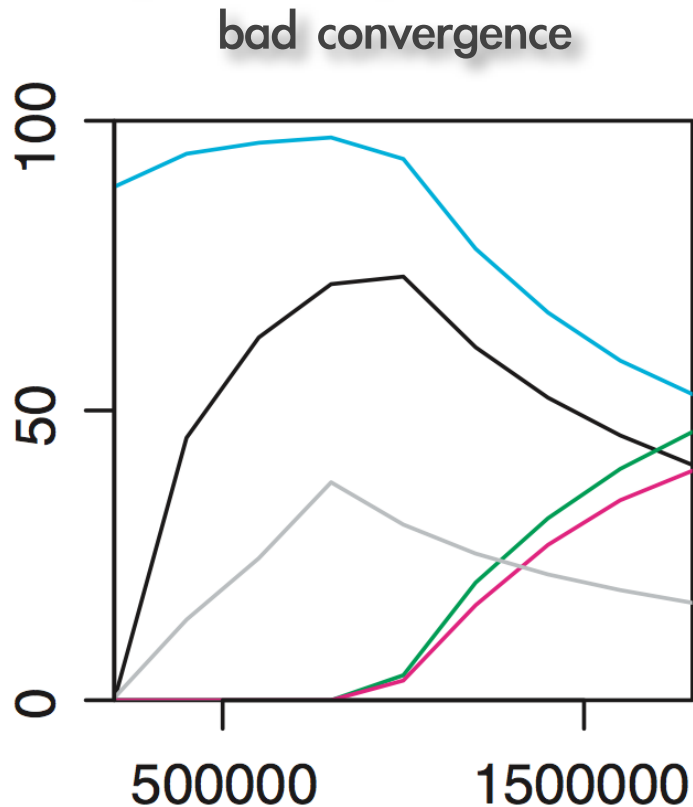
1. Convergence diagnostics

(i) Time-series plots of parameter estimates

- continuous parameters (e.g., substitution rates)--Tracer
 - some parameters are more reliable than others
 - steps may occur!
- discrete parameters (e.g., cumulative bi-partition frequency)--AWTY

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: AWTY plots of cumulative bi-partition frequency of 5 nodes



Assessing MCMC Performance: Diagnostics Based on Single Runs

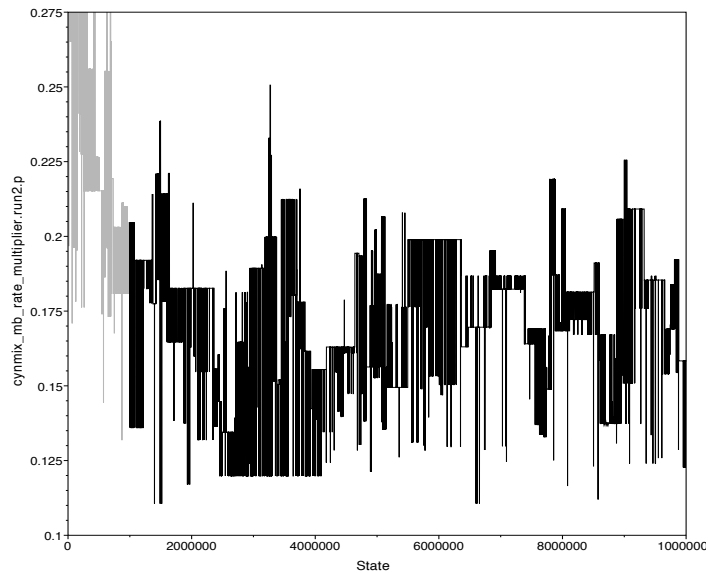
2. Mixing diagnostics

- (i) Form of the time-series plots of parameter estimates
 - continuous parameters (e.g., substitution rates)--Tracer
warm and fuzzy caterpillars

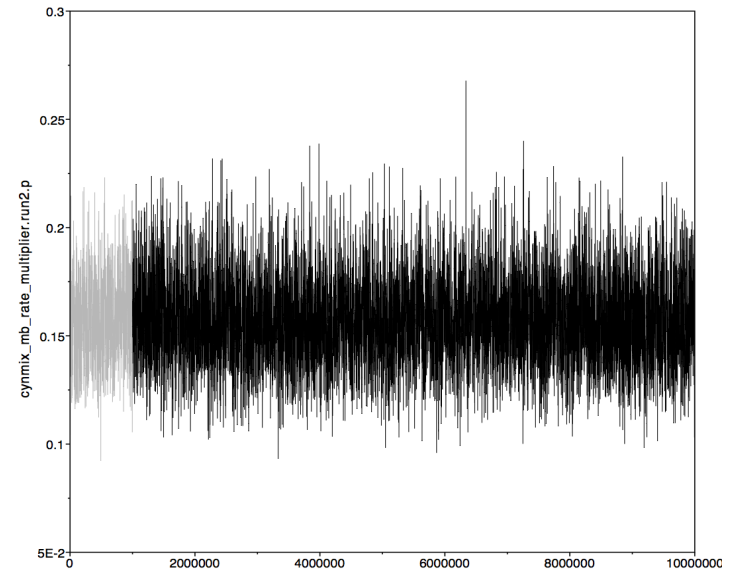
Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Tracer plots of relative-rate multipliers from two MrBayes runs

bad mixing



better mixing



Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Tracer plots of relative-rate multipliers from two MrBayes runs

bad mixing



better mixing



Assessing MCMC Performance: Diagnostics Based on Single Runs

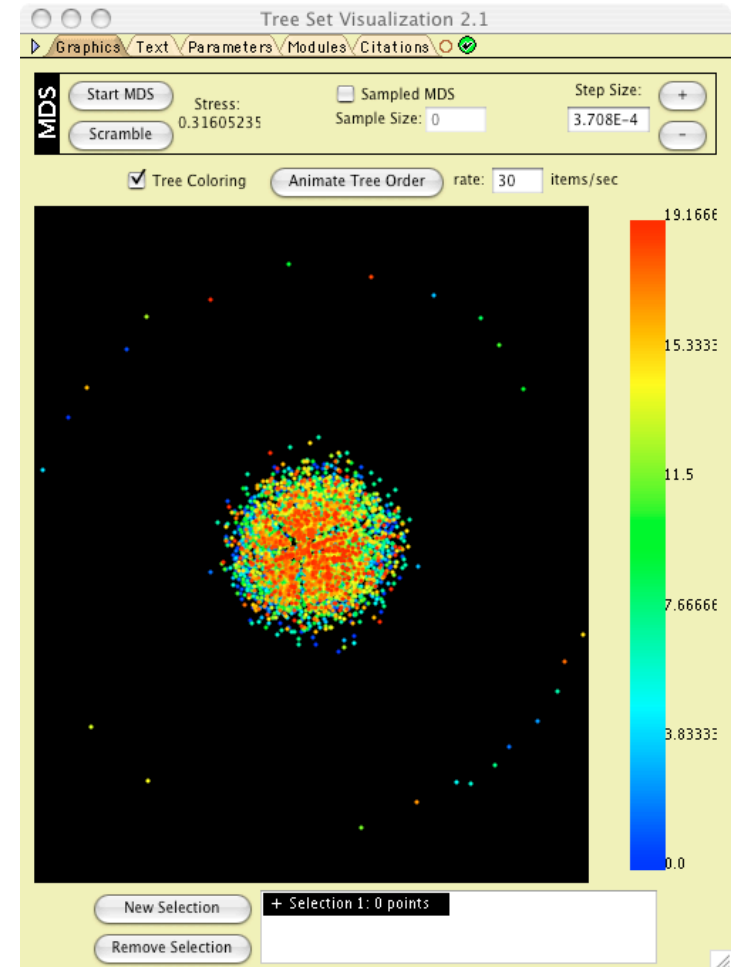
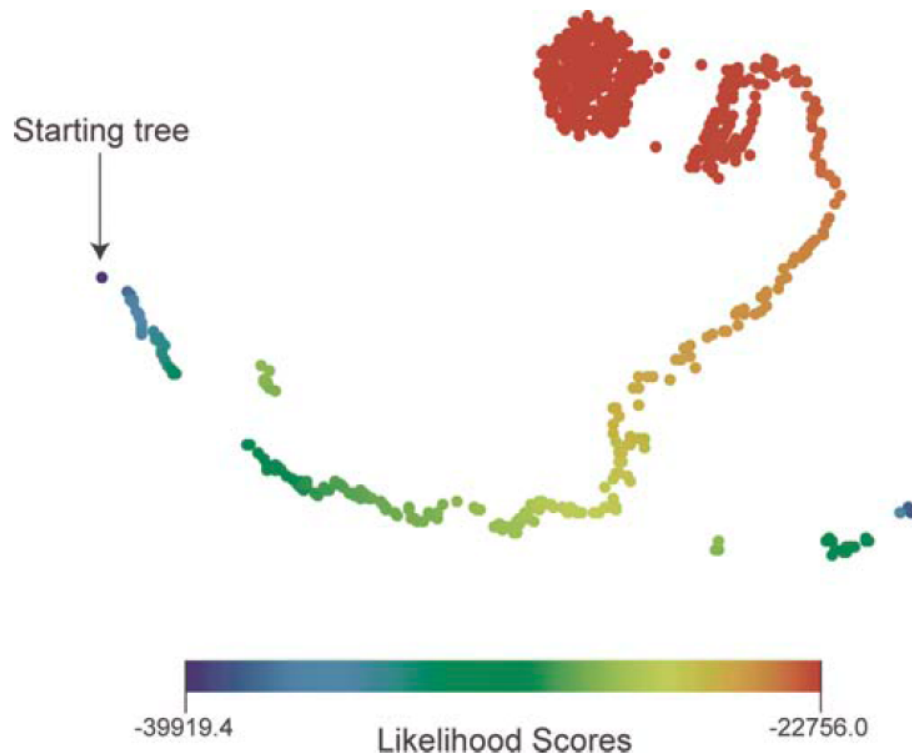
2. Mixing diagnostics

(i) Form of the time-series plots of parameter estimates

- continuous parameters (e.g., substitution rates)--Tracer
warm and fuzzy caterpillars
- discrete parameters:
 - distances among sampled topologies--TreeSetViz

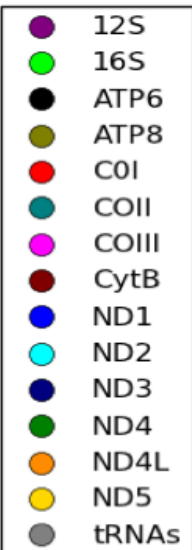
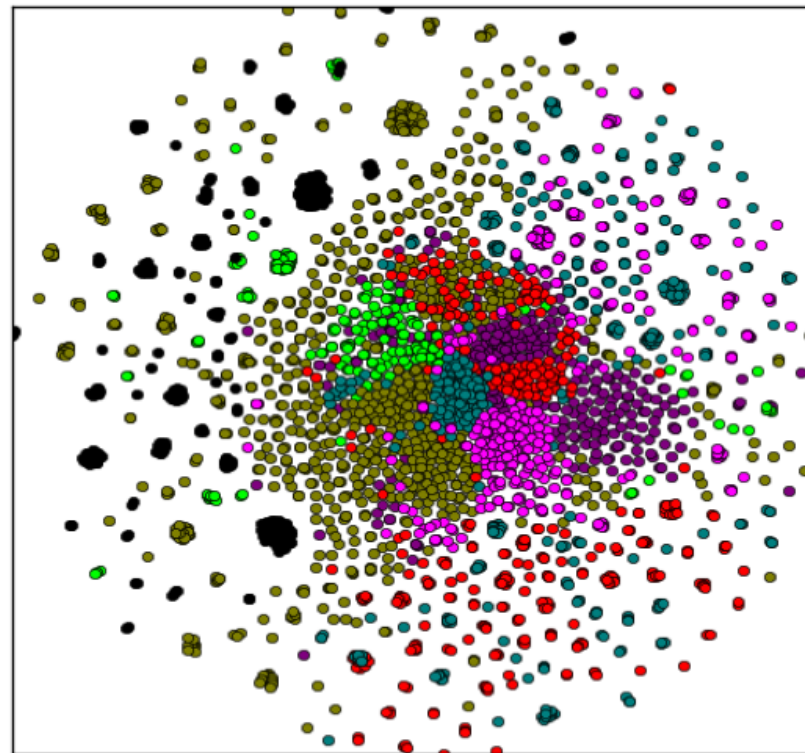
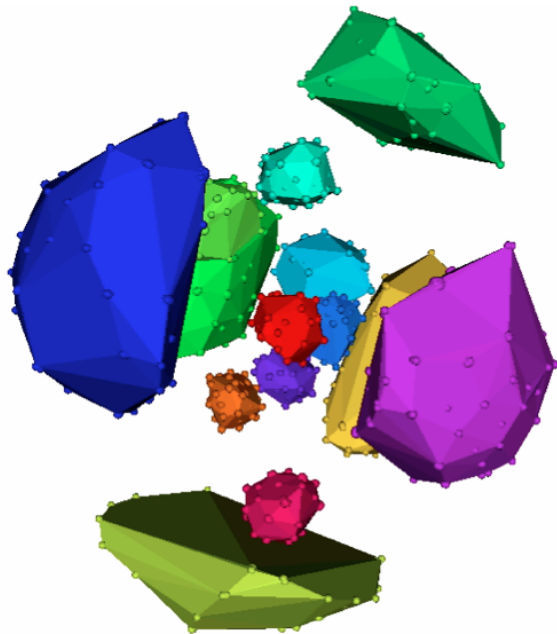
Assessing MCMC Performance: Diagnostics Based on Single Runs

TreeSetViz



Assessing MCMC Performance: Diagnostics Based on Single Runs

TreeScaper



Assessing MCMC Performance: Diagnostics Based on Single Runs

2. Mixing diagnostics

- (i) Form of the time-series plots of parameter estimates
 - continuous parameters (e.g., substitution rates)--Tracer
warm and fuzzy caterpillars
 - discrete parameters:
 - distances among sampled topologies--TreeSetViz
- (ii) Acceptance rates of parameter updates
 - continuous & discrete parameters--MrBayes, BEAST, etc.
rates should ideally fall in the ~20-70% range

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Tracer plots of relative-rate multipliers from two MrBayes runs

bad mixing



Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
13.61 % param. 1 (revmat) with Dirichlet proposal

⋮

0.04 % param. 34 (rate multiplier) Dirichlet proposal
6.59 % param. 35 (topology and branch lengths) TBR
14.06 % param. 35 (topology and branch lengths) LOCAL

better mixing



Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
33.30 % param. 1 (revmat) with Dirichlet proposal

⋮

19.13 % param. 34 (rate multiplier) Dirichlet proposal
17.40 % param. 35 (topology and branch lengths) TBR
29.76 % param. 35 (topology and branch lengths) LOCAL

Assessing MCMC Performance: Diagnostics Based on Single Runs

2. Mixing diagnostics

(i) Form of the time-series plots of parameter estimates

- continuous parameters (e.g., substitution rates)--Tracer
warm and fuzzy caterpillars
- discrete parameters:
 - distances among sampled topologies--TreeSetViz

(ii) Acceptance rates of parameter updates

- continuous & discrete parameters--MrBayes, BEAST, etc.
rates should ideally fall in the ~20-70% range
- acceptance rates can be controlled by varying the scale of the
tuning parameters for the relevant proposal mechanisms
to increase rates, decrease scale & vice versa

Assessing MCMC Performance: Diagnostics Based on Single Runs

2. Mixing diagnostics

(i) Form of the time-series plots of parameter estimates

- continuous parameters (e.g., substitution rates)--Tracer
warm and fuzzy caterpillars
- discrete parameters:
 - distances among sampled topologies--TreeSetViz

(ii) Acceptance rates of parameter updates

- continuous & discrete parameters--MrBayes, BEAST, etc.
rates should ideally fall in the ~20-70% range
- acceptance rates can be controlled by varying the scale of the
tuning parameters for the relevant proposal mechanisms
to increase rates, decrease scale & vice versa

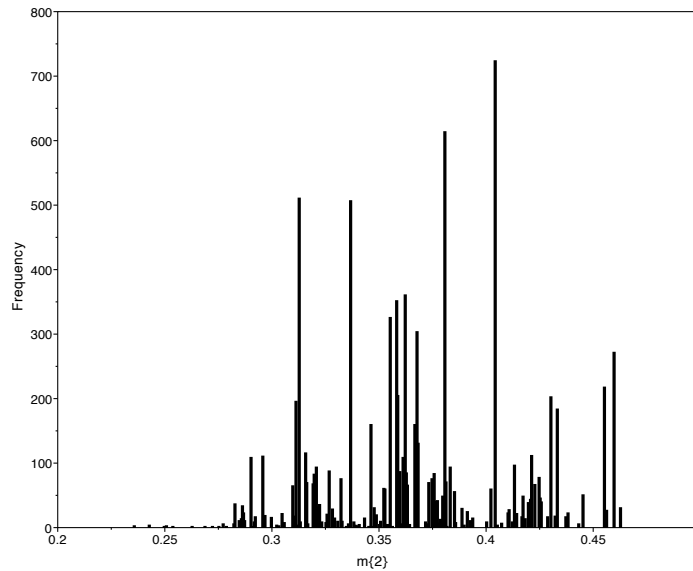
(iii) Form of the marginal posterior probability densities

- continuous parameters (e.g., substitution rates)--Tracer
beware of porcupine roadkill

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Parameter estimates for relative-rate multipliers from two MrBayes runs

bad mixing



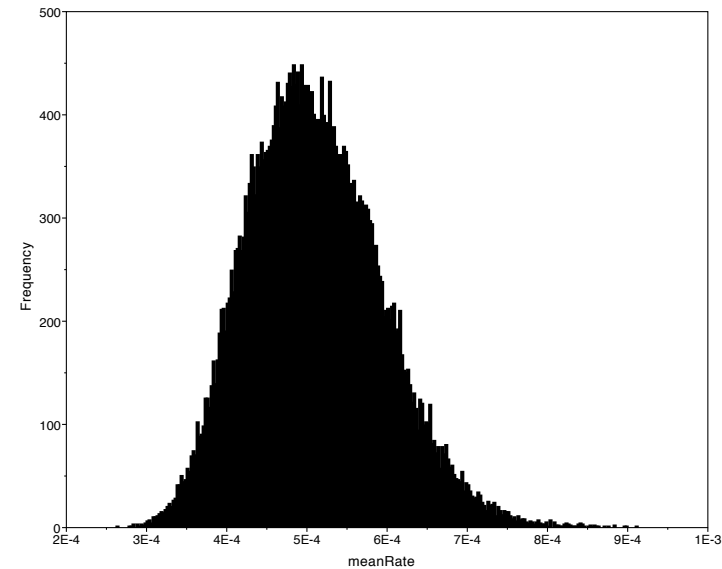
Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
13.61 % param. 1 (revmat) with Dirichlet proposal

⋮

0.04 % param. 34 (rate multiplier) Dirichlet proposal
6.59 % param. 35 (topology and branch lengths) TBR
14.06 % param. 35 (topology and branch lengths) LOCAL

better mixing



Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
33.30 % param. 1 (revmat) with Dirichlet proposal

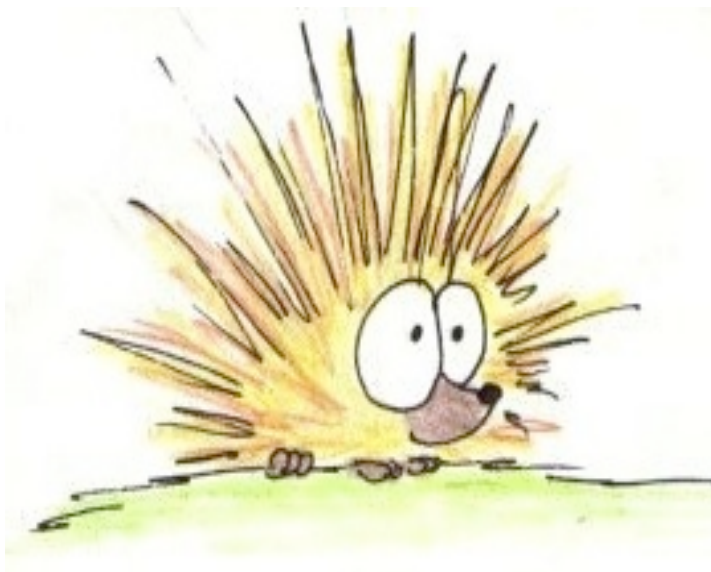
⋮

19.13 % param. 34 (rate multiplier) Dirichlet proposal
17.40 % param. 35 (topology and branch lengths) TBR
29.76 % param. 35 (topology and branch lengths) LOCAL

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Parameter estimates for relative-rate multipliers from two MrBayes runs

bad mixing



Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
13.61 % param. 1 (revmat) with Dirichlet proposal

.

.

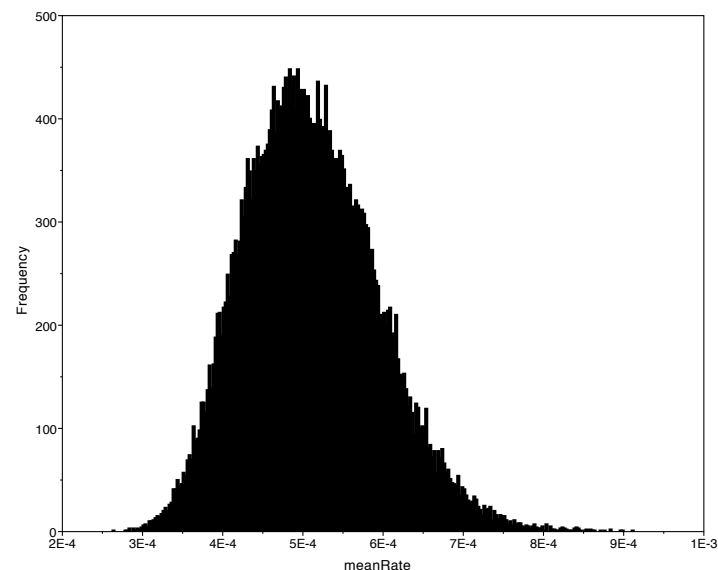
.

0.04 % param. 34 (rate multiplier) Dirichlet proposal

6.59 % param. 35 (topology and branch lengths) TBR

14.06 % param. 35 (topology and branch lengths) LOCAL

better mixing



Acceptance rates for the moves in the "cold" chain of run 1:

With prob. Chain accepted changes to
33.30 % param. 1 (revmat) with Dirichlet proposal

.

.

.

19.13 % param. 34 (rate multiplier) Dirichlet proposal

17.40 % param. 35 (topology and branch lengths) TBR

29.76 % param. 35 (topology and branch lengths) LOCAL

Assessing MCMC Performance: Diagnostics Based on Single Runs

2. Mixing diagnostics

(iv) Autocorrelation time (ACT) of parameter samples

- The lag k autocorrelation ρ_k is the correlation every draw and its k th lag:

$$\rho_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

We would expect the k th lag autocorrelation to be smaller as k increases (our 1st and 100th draws should be less correlated than our 1st and 2nd draws).

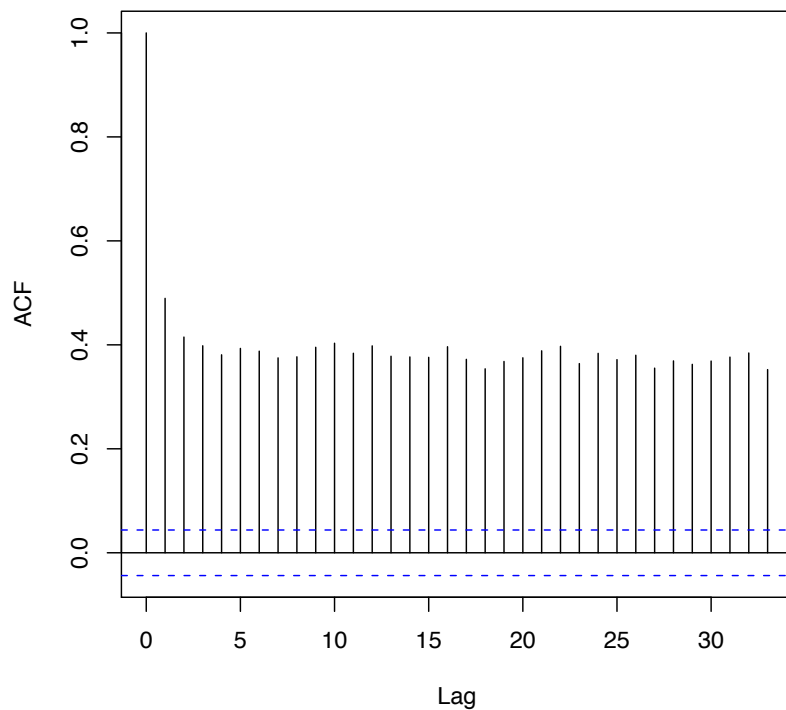
If autocorrelation is still relatively high for higher values of k , this indicates high degree of correlation between our draws and slow mixing.

Assessing MCMC Performance: Diagnostics Based on Single Runs

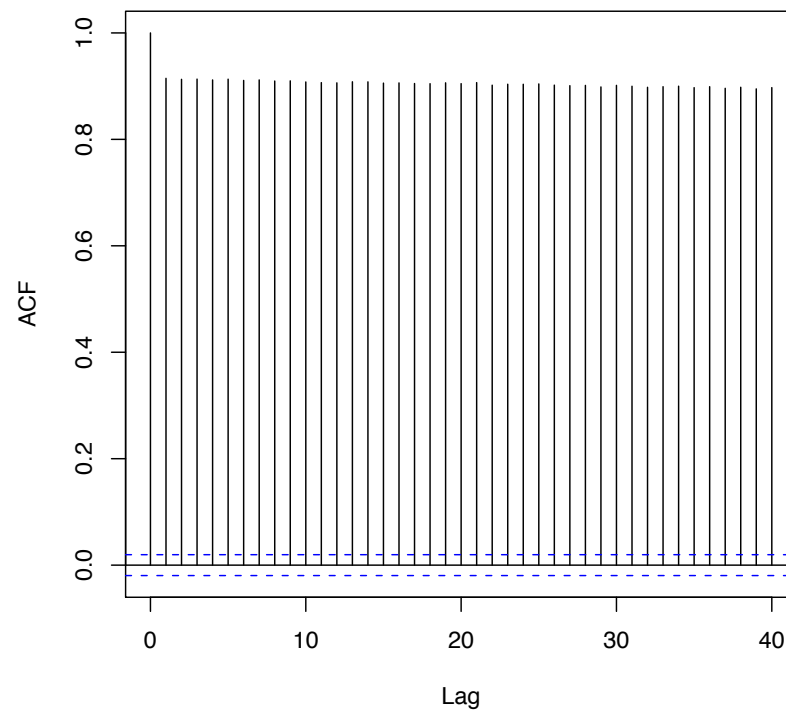
2. Mixing diagnostics

(iv) Autocorrelation time (ACT) of parameter samples

better mixing



bad mixing



Assessing MCMC Performance: Diagnostics Based on Single Runs

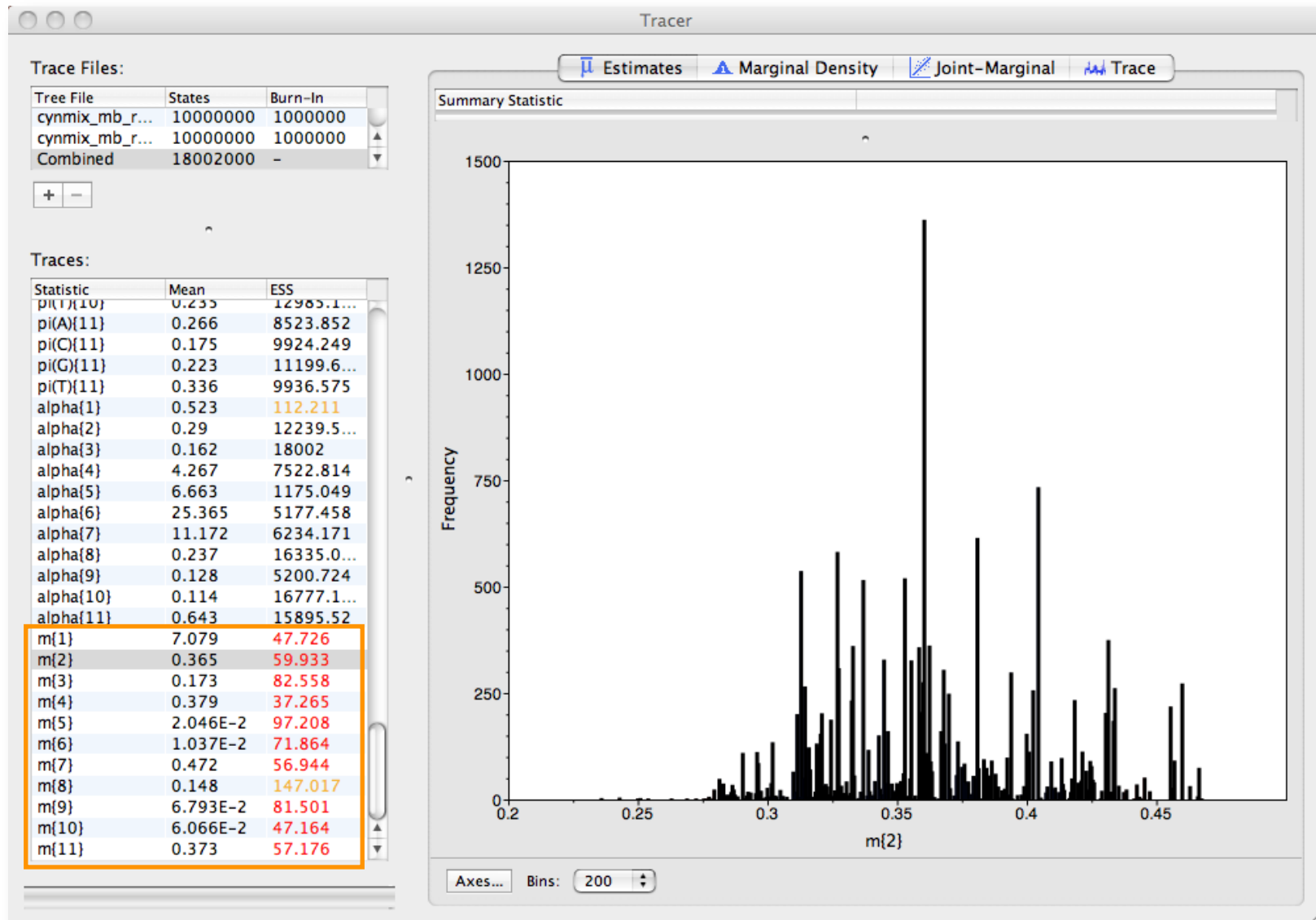
3. Sample-size diagnostics

(i) Effective Sample Size (ESS) diagnostic

- number of samples/autocorrelation time (ACT)
- continuous parameters (e.g., substitution rates)--Tracer

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: ESS values for relative-rate multipliers from two MrBayes runs
low intensity



Assessing MCMC Performance: Diagnostics Based on Single Runs

3. Sample-size diagnostics

(i) Effective Sample Size (ESS) diagnostic

- number of samples/autocorrelation time (ACT)
- continuous parameters (e.g., substitution rates)--Tracer

(ii) Form of the marginal posterior probability densities

- continuous parameters (e.g., substitution rates)--Tracer
brother of porcupine roadkill
ensure SAE compliance!

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Parameter estimates for mean-rate multipliers from BEAST runs

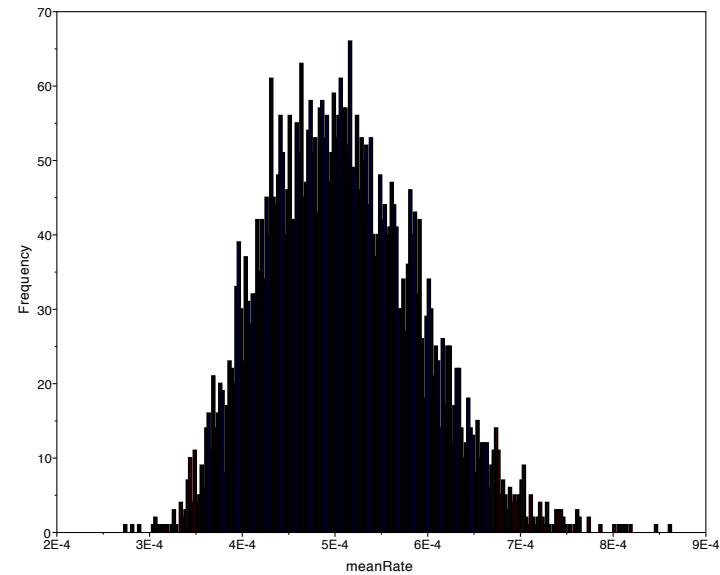
low intensity



1M cycles

- inadequate chain length

better intensity



5M cycles

Assessing MCMC Performance: Diagnostics Based on Single Runs

Example: Parameter estimates for mean-rate multipliers from BEAST runs

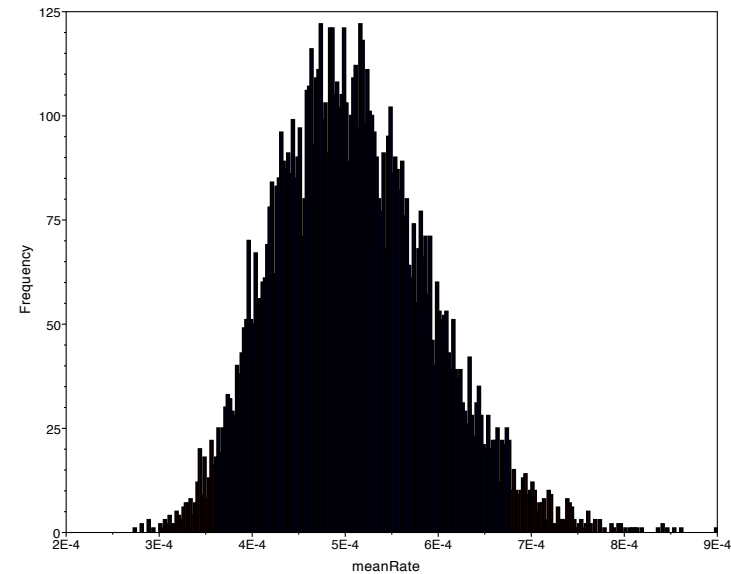
low intensity



1M cycles

- inadequate chain length

better intensity



10M cycles

Assessing MCMC Performance: Diagnostics Based on Single Runs

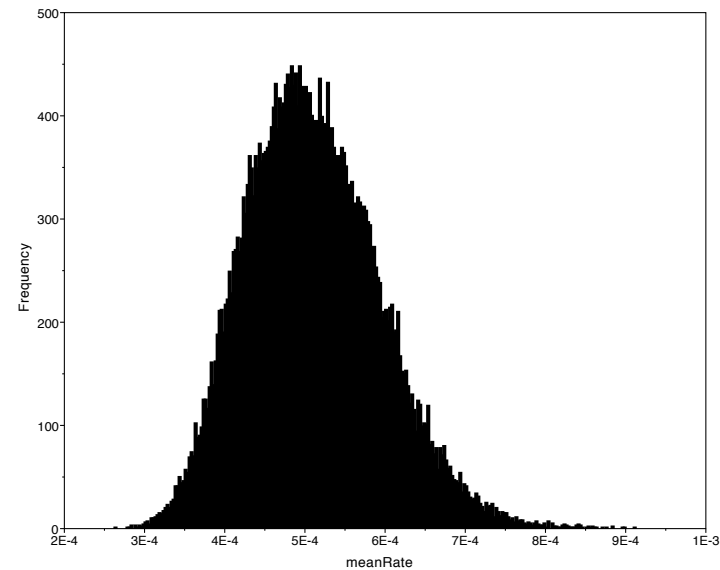
Example: Parameter estimates for relative-rate multipliers from two MrBayes runs

low intensity



1M cycles

better intensity

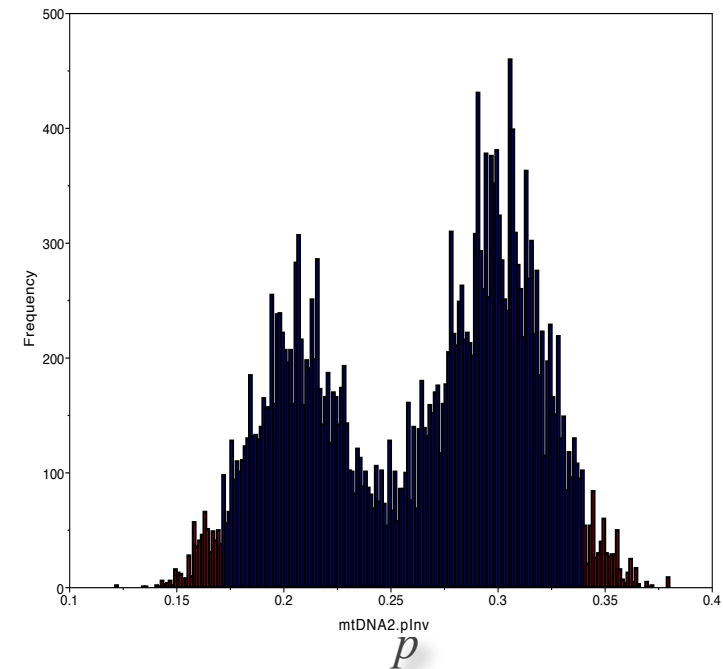
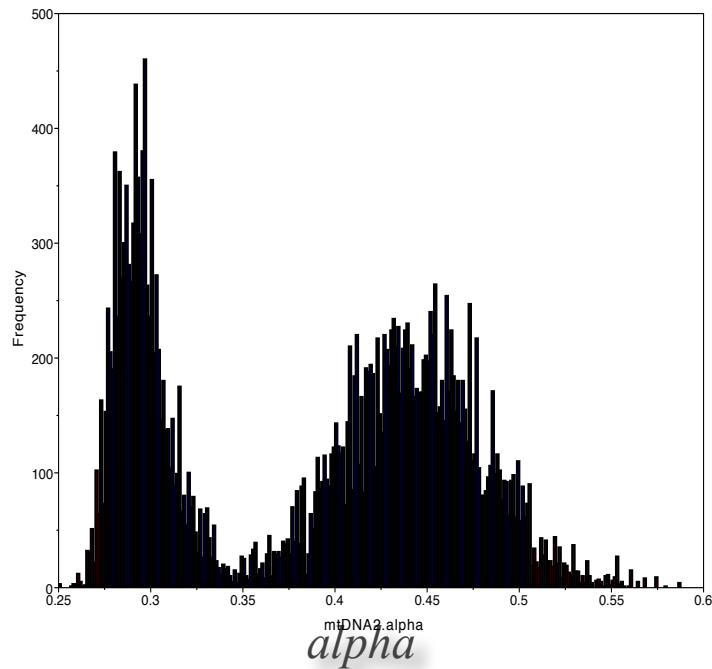


40M cycles

Assessing MCMC Performance: Diagnostics Based on Single Runs

MCMC pathologies

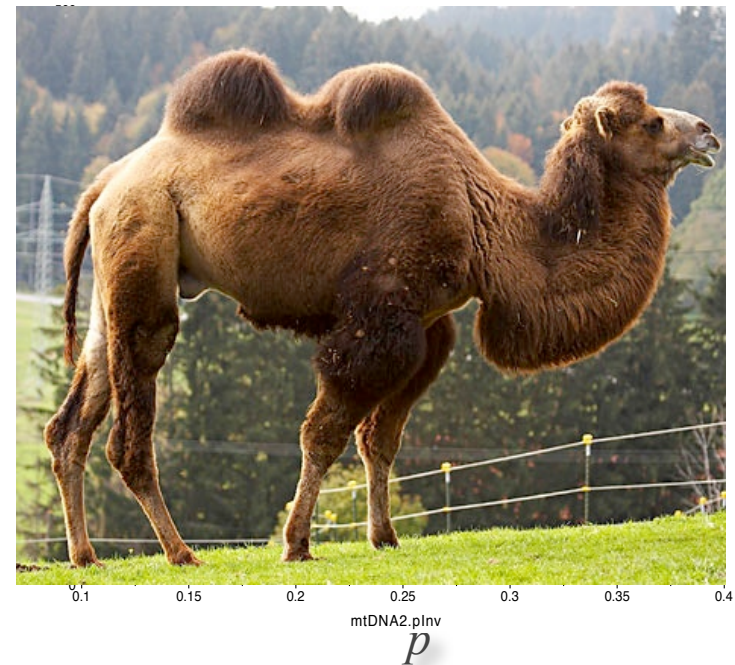
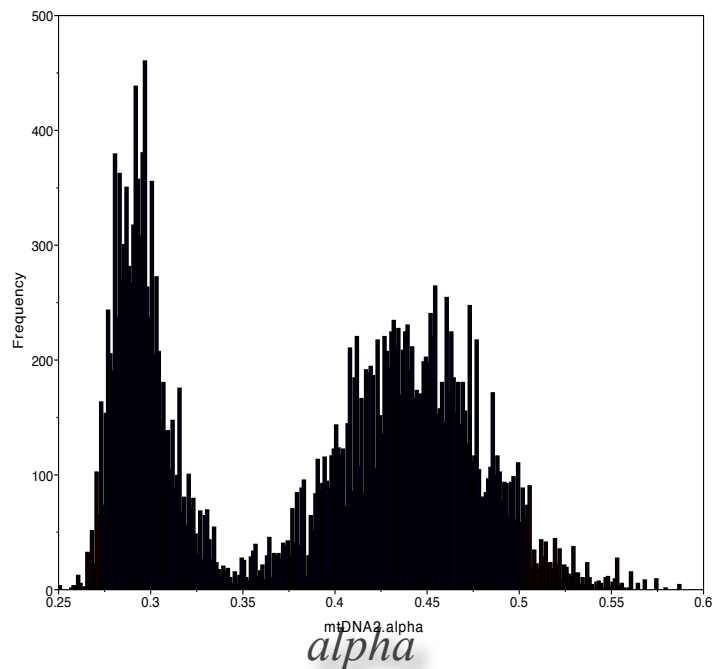
Parameter interaction between I+G mixture for among-site rate variation



Assessing MCMC Performance: Diagnostics Based on Single Runs

MCMC pathologies

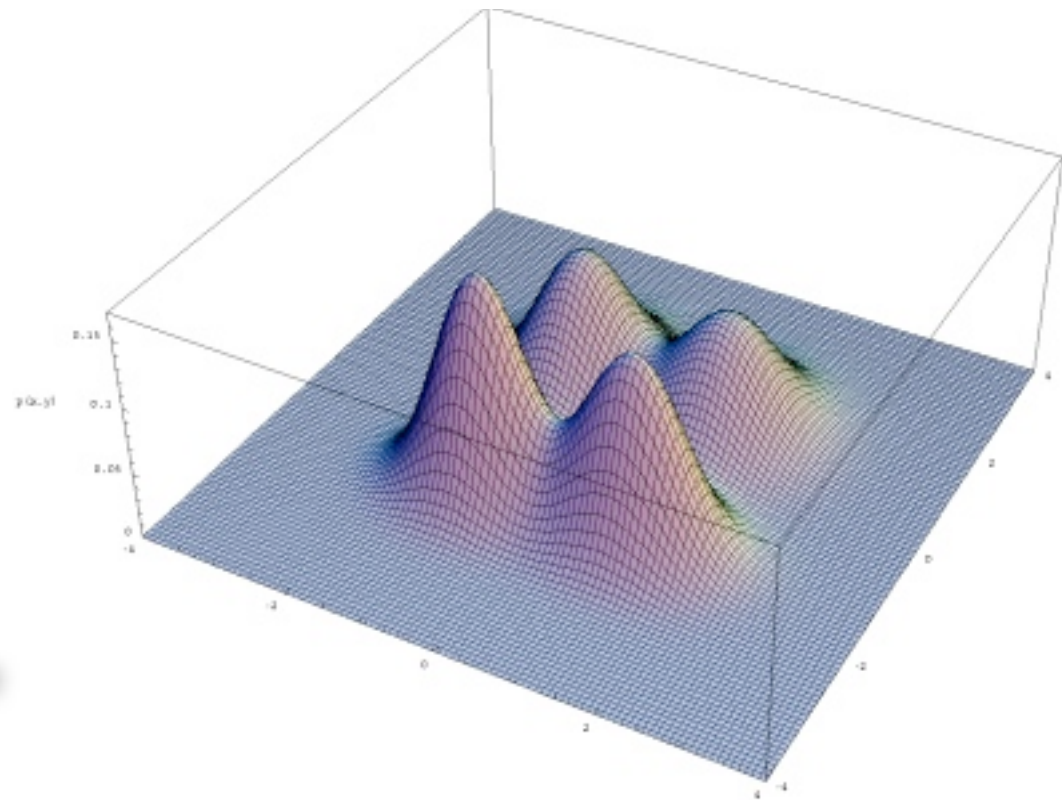
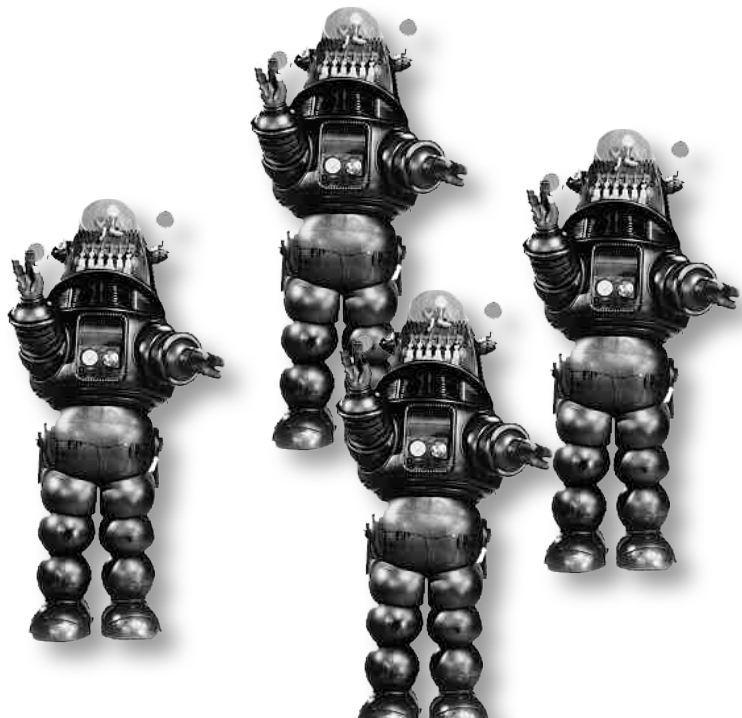
Parameter interaction between I+G mixture for among-site rate variation



- multi-modal marginal densities stem from non-identifiability
- use G with additional discrete rate categories

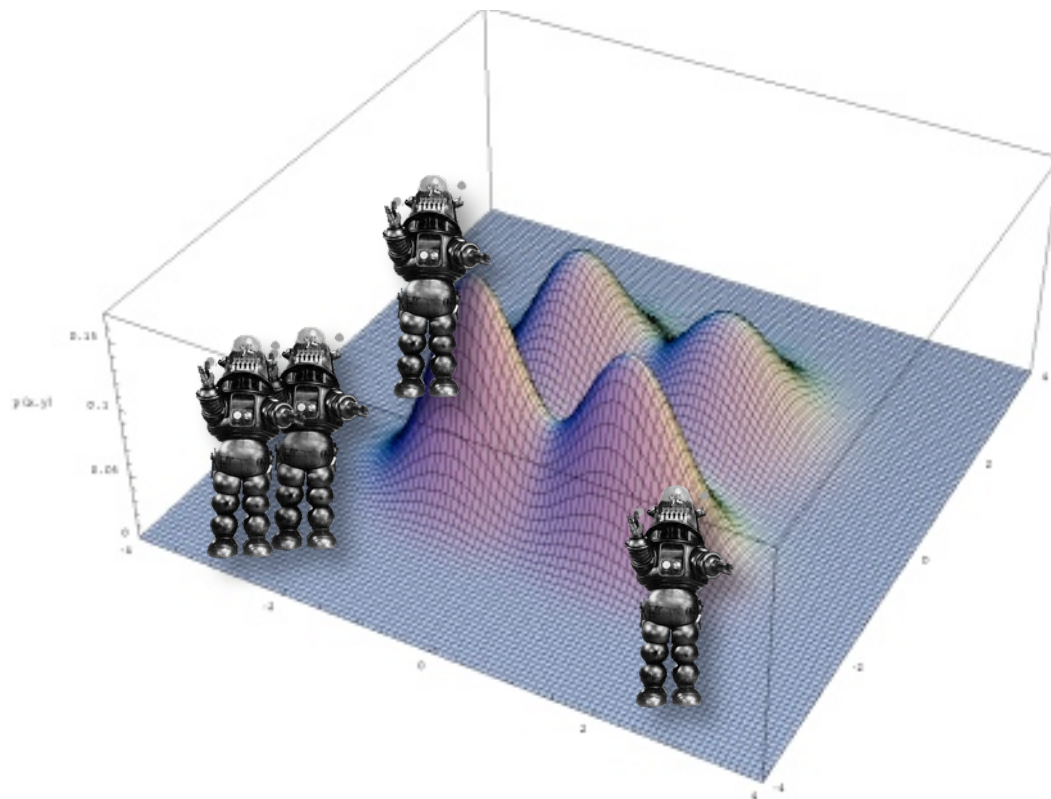
Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

Robot Squadron!!



Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

Robot Squadron!!



Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

A slightly more formal description...

To facilitate mixing over the joint posterior probability density, multiple incrementally heated chains may be run

N chains are initiated from random starting point in the joint posterior probability density.

One chain is cold, and $N-1$ are incrementally heated.

Samples are drawn from the cold chain.

The heating distorts the joint posterior probability density, such that chains can more freely traverse regions of the stationary distribution.

Occasionally, a swap is attempted between the cold and one of the randomly chosen heated chains, which ensures that samples are drawn from regions of high posterior probability.

heat of chain $i = 1/(1 + iT)$

chain	0.25	0.20	0.15	0.10
1	1.00	1.00	1.00	1.00
2	0.80	0.83	0.87	0.91
3	0.66	0.71	0.77	0.83
4	0.57	0.62	0.69	0.77

Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

Diagnosing MC³ performance

The primary diagnostic is the acceptance rates for proposed chain swaps:

As a rule of thumb, acceptance rates for proposals should fall in ~20–70% range

- if acceptance rates are too low, decrease the value of the temperature parameter
- if acceptance rates are too high, increase the value of the temperature parameter

Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

Example: Tracer plots of relative-rate multipliers from two MrBayes runs

bad mixing

better mixing

Chain swap information for run 1:

	1	2	3	4
1		0.01	0.00	0.00
2	1666381		0.01	0.00
3	1666964	1664302		0.20
4	1666923	1668351	1667079	

Chain swap information for run 2:

	1	2	3	4
1		0.00	0.00	0.00
2	1664180		0.16	0.00
3	1667247	1669245		0.04
4	1665043	1667632	1666653	

Chain swap information for run 1:

	1	2	3	4
1		0.60	0.32	0.17
2	834663		0.65	0.40
3	832631	834125		0.70
4	831509	834020	833052	

Chain swap information for run 2:

	1	2	3	4
1		0.60	0.32	0.17
2	833614		0.65	0.40
3	834623	833715		0.70
4	833536	832594	831918	

Approximating the Joint Posterior Probability Density with Metropolis-Couples MCMC

Diagnosing MC³ performance

The primary diagnostic is the acceptance rates for proposed chain swaps:

As a rule of thumb, acceptance rates for proposals should fall in ~20–70% range

- if acceptance rates are too low, decrease the value of the temperature parameter
- if acceptance rates are too high, increase the value of the temperature parameter

Other aspects controlling the behavior of the Metropolis coupling can be modified to improve MCMC performance:

- increase the number of incrementally heated chains (e.g., `nchains` parameter)
- increase the frequency of attempted chain-swap events (e.g., `swapfreq` parameter)
- increase the number of swaps attempted per event (e.g., `nswaps` parameter)

Outline

- I. A review of where we've been and why
- II. Diagnosing MCMC performance
- III. Diagnostics based on single chains
-  IV. Diagnostics based on the prior
- V. Diagnostics based on multiple chains

Assessing MCMC Performance: Diagnostics Based on the Prior

Estimating under the prior...

Marginal posterior densities for parameters are updated versions of the corresponding prior probability densities: they are updated by the information in the data via the likelihood function

$$\text{posterior probability } \Pr[\tau_i | X] = \frac{\text{likelihood function } \Pr[X | \tau_i] \times \text{prior probability } \Pr[\tau_i]}{\text{marginal likelihood } \sum_{j=1}^{B(s)} \Pr[X | \tau_j] \times \Pr[\tau_j]}$$

Assessing MCMC Performance: Diagnostics Based on the Prior

Estimating under the prior...

Marginal posterior densities for parameters are updated versions of the corresponding prior probability densities: they are updated by the information in the data via the likelihood function

We can compare the marginal prior densities to their posterior counterparts to help identify weak parameters

- MCMC can be run to target the joint prior either by estimating with no data or by forcing the likelihood function return 1.

$$R = \min \left[1, \frac{f(X|\Theta')}{f(X|\Theta)} \cdot \frac{f(\Theta')}{f(\Theta)} \cdot \frac{f(\Theta|\Theta')}{f(\Theta'|\Theta)} \right]$$

likelihood ratio prior ratio proposal ratio

Outline

I. A review of where we've been and why

II. Diagnosing MCMC performance

III. Diagnostics based on single chains

IV. Diagnostics based on the prior

 V. Diagnostics based on multiple chains

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

The general idea is to compare estimates from multiple independent chains initiated from random parameter values

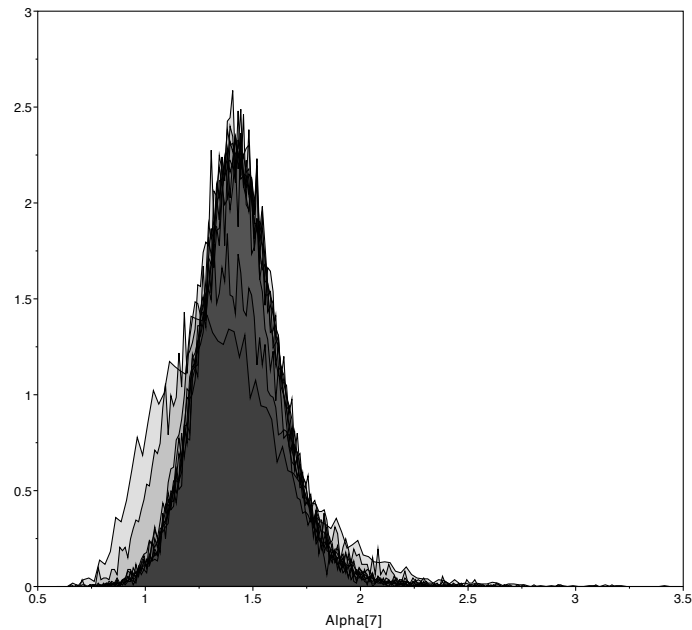
Form of the marginal posterior densities for all parameters

- continuous parameters (e.g., substitution rates)--Tracer

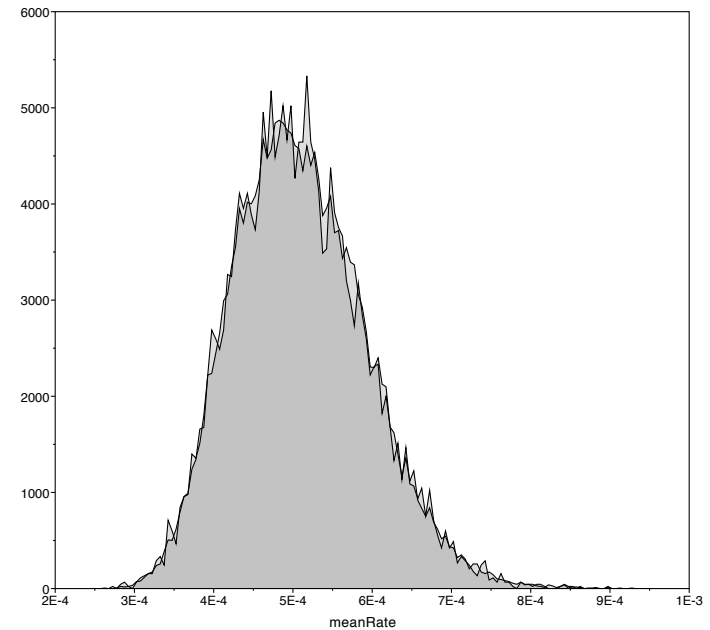
Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: Tracer plots of marginal densities from multiple MrBayes runs

bad convergence



better convergence



*Tracer demo

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

The general idea is to compare estimates from multiple independent chains initiated from random parameter values

Form of the marginal posterior densities for all parameter

- continuous parameters:

- PSRF diagnostic--MrBayes

1. Run $m \geq 2$ chains of length $2n$ from overdispersed starting values.
2. Discard the first n draws of each chain.
3. Calculate the within-chain and between-chain variance.
4. Calculate the estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance.
5. Calculate the PSRF.

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: PSRF values for relative-rate multipliers from two MrBayes runs

bad convergence

Parameter	Mean	Variance	95% Cred. Interval		Median	PSRF *
			Lower	Upper		
TL{all}	4.921609	2.998138	2.836000	7.295000	5.056000	9.084
kappa{4,5}	3.095696	0.054125	2.667623	3.587024	3.085271	1.000
alpha{5}	1.006544	0.087721	0.606472	1.738482	0.950093	1.000
pinvar{1}	0.307396	0.009357	0.095913	0.471070	0.316173	1.000
m{1}	0.264226	0.009315	0.146502	0.421870	0.244468	5.507
m{2}	0.040919	0.000227	0.022205	0.065884	0.037425	5.279
m{3}	2.721453	7.157157	0.039001	5.544253	5.030560	69.564
m{4}	2.125810	3.568002	0.199137	4.044249	3.917338	150.012
m{5}	0.188768	0.004373	0.109303	0.295129	0.170624	5.749

better convergence

Parameter	Mean	Variance	95% Cred. Interval		Median	PSRF *
			Lower	Upper		
TL{all}	0.073893	0.000034	0.063000	0.086000	0.074000	1.000
kappa{2,3}	3.236308	0.366904	2.199024	4.587719	3.190195	1.000
m{1}	1.285838	0.028345	0.980634	1.630387	1.278161	1.000
m{2}	1.423906	0.015507	1.182596	1.664627	1.423610	1.000
m{3}	0.589346	0.005341	0.453175	0.736459	0.587617	1.001

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

The general idea is to compare estimates from multiple independent chains initiated from random parameter values

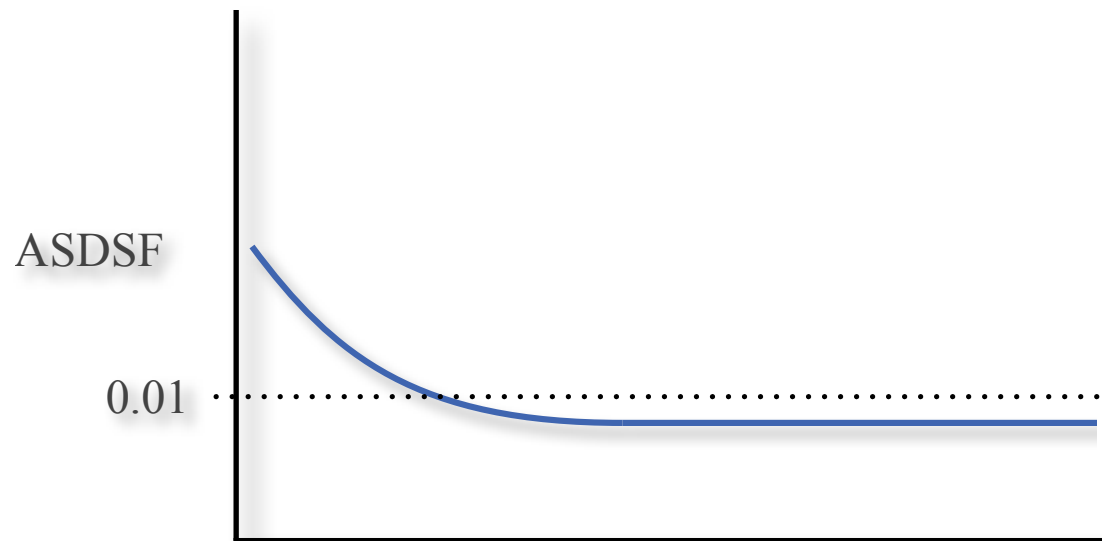
Form of the marginal posterior densities for all parameter

- continuous parameters:
 - similarity of marginal densities--Tracer
 - PSRF diagnostic--MrBayes
- discrete parameters:
 - Topology
 - similarity of paired chains (e.g., ASDSF diagnostic in MrBayes)

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

MCMC pathologies

Time-series plot of ASDSF diagnostic

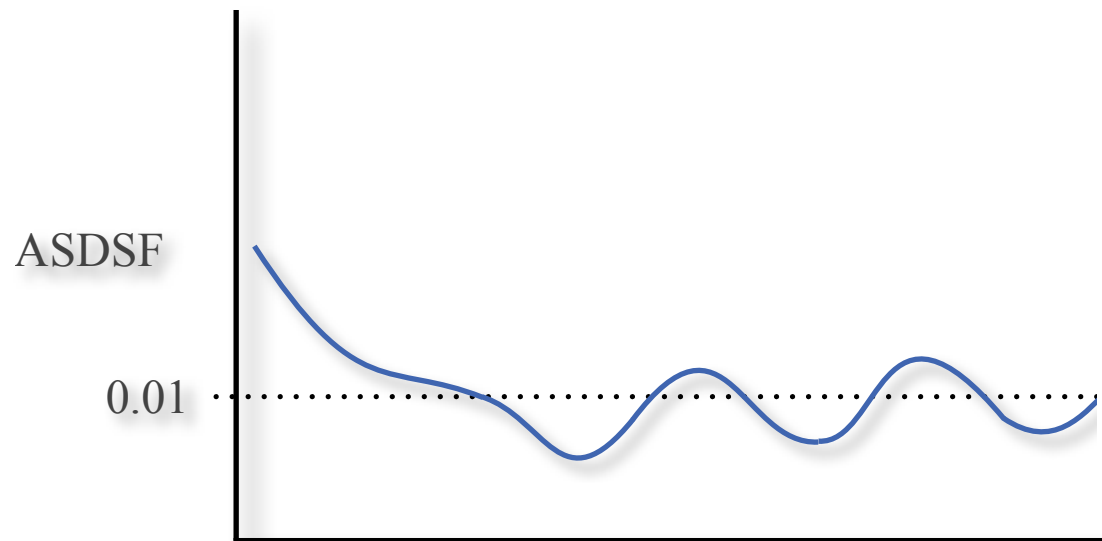


- stop sampling when $ASDSF < 0.01$
- under-parameterized model

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

MCMC pathologies

Time-series plot of ASDSF diagnostic



- ASDSF oscillate about the 0.01 threshold value
- over-parameterized model
- poorly chosen priors

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

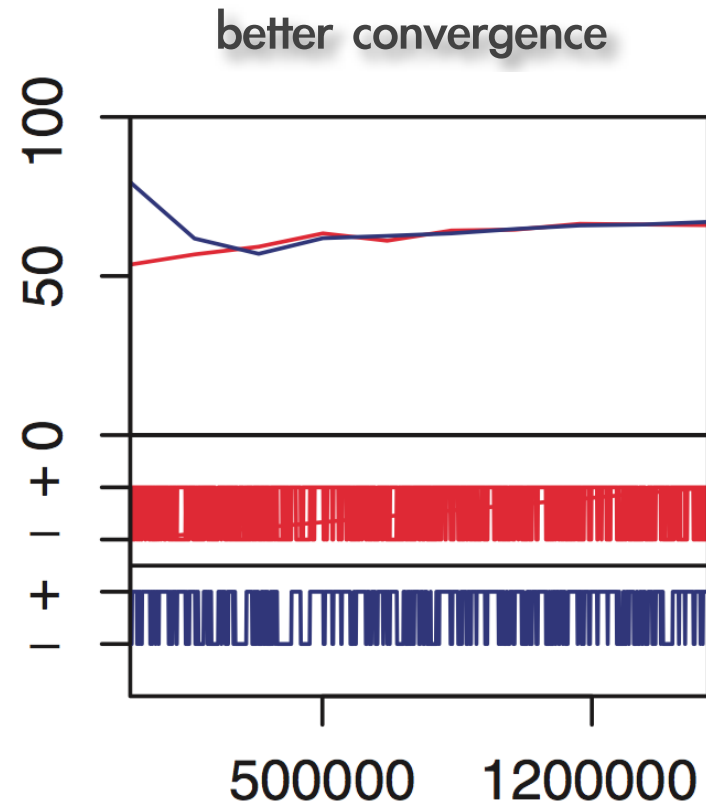
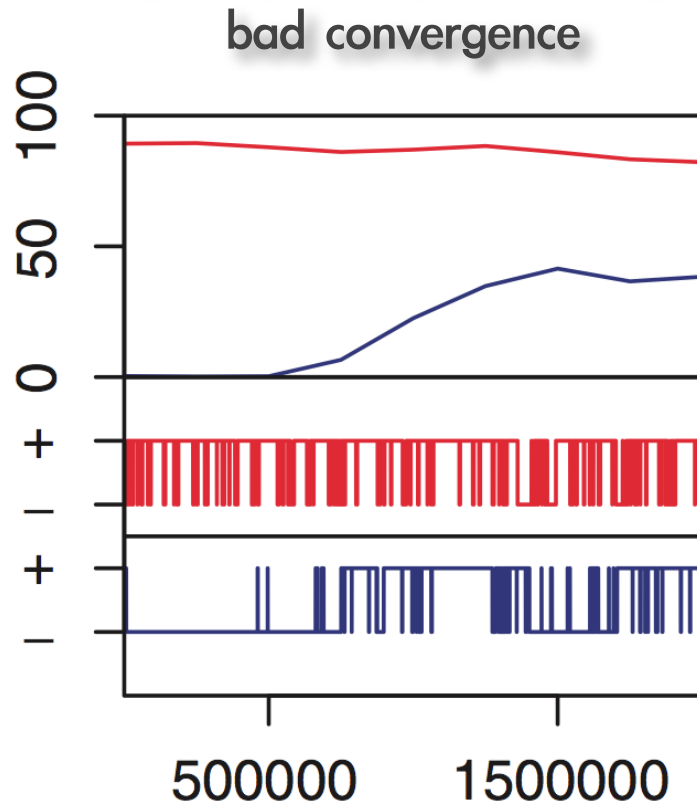
The general idea is to compare estimates from multiple independent chains initiated from random parameter values

Form of the marginal posterior densities for all parameter

- continuous parameters (e.g., substitution rates)--Tracer
- discrete parameters:
 - Topology
 - similarity of paired chains (e.g., ASDSF diagnostic in MrBayes)
 - distances among sampled topologies--TreeSetViz/TreeScaper
 - split frequencies & presence/absence--AWTY

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: split frequencies & presence/absence in AWTY



Assessing MCMC Performance: Diagnostics Based on Multiple Runs

The general idea is to compare estimates from multiple independent chains initiated from random parameter values

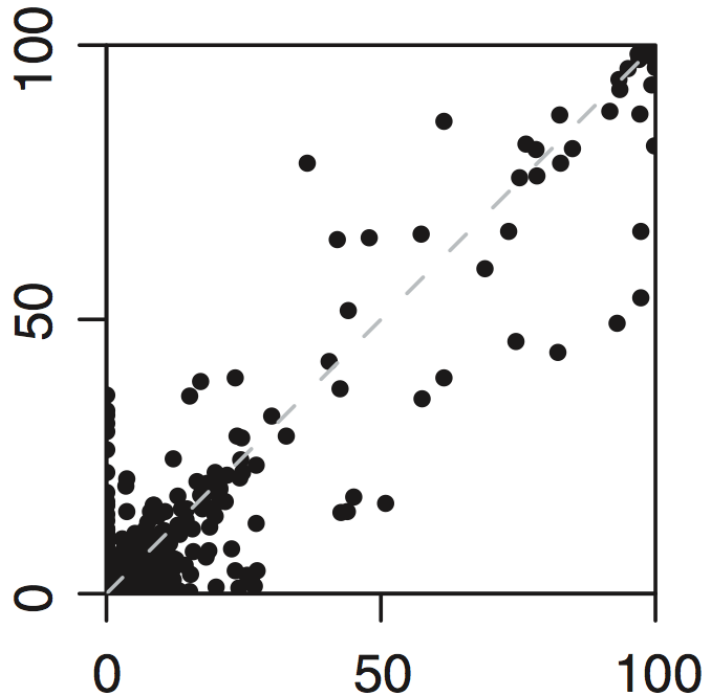
Form of the marginal posterior densities for all parameter

- continuous parameters (e.g., substitution rates)--Tracer
- discrete parameters:
 - Topology
 - similarity of paired chains (e.g., ASDSF diagnostic in MrBayes)
 - distances among sampled topologies--TreeSetViz/TreeScaper
 - split frequencies & presence/absence--AWTY/BPD
 - nodal support--AWTY/MrBayes

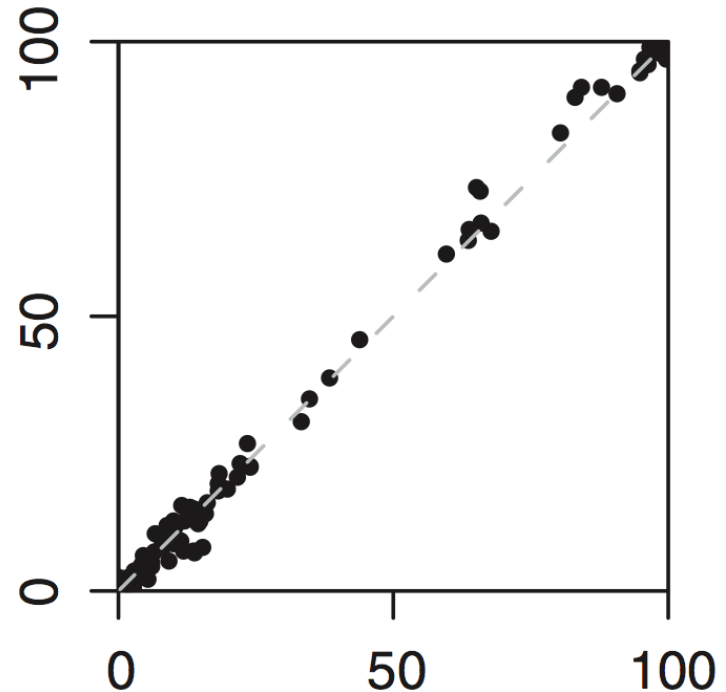
Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: ‘comparetrees’ plot of trees sampled by two MrBayes runs

bad convergence



better convergence



Summary: Some General Strategies for Assessing MCMC Performance

You can never be absolutely certain that the MCMC is reliable, you can only identify when something has gone wrong. Gelman

1. When do you need to assess MCMC performance?

ALWAYS

2. When should you assess the performance of individual runs?

ALWAYS

3. Which diagnostics should you use to assess individual runs?

ALL that are relevant for the models/parameters you are estimating under

4. When is a single run sufficient to assess MCMC performance?

NEVER

5. When should you estimate under the prior?

WHENEVER POSSIBLE (and be wary of programs where it is not possible)

Summary: Some General Strategies for Assessing MCMC Performance

You can never be absolutely certain that the MCMC is reliable, you can only identify when something has gone wrong. Gelman

6. When should you use Metropolis-Coupling?

Whenever you cannot be certain that standard MCMC is adequate
i.e., **ALWAYS** (and be wary of programs where it is not possible)

7. When should you perform multiple independent MCMC runs?

ALWAYS (and be wary of pseudo-independence)

8. Which diagnostics should you use to assess individual runs?

ALL that are relevant for the models/parameters you are estimating under

9. How many independent MCMC runs are sufficient?

AS MANY AS POSSIBLE (i.e., as many as you think your data/problem deserve)

10. How long should you run each MCMC analysis?

AS LONG AS POSSIBLE (i.e., as long as you think your data/problem deserve)

Assessing MCMC Performance: Software Tools

Tracer

AWTY (BPD)

TreeSetViz

TreeScaper

BOA

coda