A Brief Introduction to Bayesian Inference of Phylogeny

Mike May
Department of Evolution & Ecology
University of California, Davis
CoME, 2022

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Statistical Estimation of Phylogeny: An Outline

Data

sequence alignment

Model

tree with branch lengths

substitution model

Inference Method

maximum likelihood

Bayesian inference

Parameter Estimates

phylogeny & other model parameters

Statistical Estimation of Phylogeny: An Outline

Data

sequence
alignment

Model

tree with
branch lengths

substitution
model

Inference Method

maximum
likelihood

Bayesian
inference

Parameter Estimates

phylogeny &
other model
parameters

Statistical Estimation of Phylogeny: An Outline

Data

sequence
alignment

Model

tree with
branch lengths

substitution
model

Inference Method

maximum
likelihood

Bayesian
inference

Parameter Estimates

phylogeny &
other model
parameters

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Bayesian Inference

Conditional Probability

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}$$

Bayesian Inference

Conditional Probability

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

Bayesian Inference

Conditional Probability

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

$$\Pr(A,B) = \Pr(B)\Pr(A \mid B)$$

and by the same reasoning:

$$\Pr(A,B) = \Pr(A)\Pr(B \mid A)$$

which is the probability of observing A times the probability of observing B given that A has occurred.

Bayesian Inference

Conditional Probability

The probability of observing A given that B has occurred, $\Pr(A \mid B)$, is just the fraction of cases in which B occurs, $\Pr(B)$, that A also occurs, $\Pr(A,B)$.

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

$$\Pr(A,B) = \Pr(B)\Pr(A \mid B)$$

and by the same reasoning:

$$\Pr(A,B) = \Pr(A)\Pr(B \mid A)$$

which is the probability of observing A times the probability of observing B given that A has occurred.

Bayesian Inference

~~Conditional Probability~~ Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\Pr(A \mid B) = \frac{\Pr(A)\Pr(B \mid A)}{\Pr(B)}$$

Joint Probability

The probability of observing both A and B, $\Pr(A,B)$, is therefore:

$$\Pr(A,B) = \Pr(B)\Pr(A \mid B)$$

and by the same reasoning:

$$\Pr(A,B) = \Pr(A)\Pr(B \mid A)$$

which is the probability of observing A times the probability of observing B given that A has occurred.

Bayesian Inference

Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\Pr(A \mid B) = \frac{\Pr(A)\Pr(B \mid A)}{\Pr(B)}$$

Bayesian Inference

Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

Bayesian Inference

Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\Pr(\boxed{A} \mid \boxed{B}) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

unobserved outcome

observed outcome

$$f(\boxed{\theta_i} \mid \boxed{\mathbf{X}}) = \frac{f(\mathbf{X} \mid \theta_i) f(\theta_i)}{\sum_{j=1}^{N} f(\mathbf{X} \mid \theta_j) f(\theta_j)}$$

Bayesian Inference

Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\boxed{\Pr(A \mid B)} = \frac{\boxed{\Pr(B \mid A)}\,\boxed{\Pr(A)}}{\Pr(B)}$$

posterior probability

likelihood function

prior probability

$$\boxed{f(\theta_i \mid \mathbf{X})} = \frac{\boxed{f(\mathbf{X} \mid \theta_i)}\,\boxed{f(\theta_i)}}{\boxed{\sum_{j=1}^{N} f(\mathbf{X} \mid \theta_j) f(\theta_j)}}$$

marginal likelihood

Bayesian Inference

Bayes Theorem

The posterior probability of observing A given that B has occurred, $\Pr(A \mid B)$, is proportional to the product of the conditional probability of $\Pr(A \mid B)$ and the unconditional probability of A, $\Pr(A)$.

$$\boxed{\Pr(A \mid B)} = \frac{\boxed{\Pr(B \mid A)}\,\boxed{\Pr(A)}}{\Pr(B)}$$

posterior probability

likelihood function

prior probability

$$\boxed{f(\theta_i \mid \mathbf{X})} = \frac{\boxed{f(\mathbf{X} \mid \theta_i)}\,\boxed{f(\theta_i)}}{\boxed{\int_\theta f(\mathbf{X} \mid \theta) f(\theta)\, d\theta}}$$

marginal likelihood

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations:

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(⚃, ⚅ \mid \text{Fair}) =$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\boxed{\frac{1}{6}}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(⚃, ⚅ \mid \text{Fair}) = \boxed{\frac{1}{6}}$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
$\boxed{\text{⚅}}$	$\boxed{\frac{1}{6}}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(\text{⚃}, \text{⚅} \mid \text{Fair}) = \frac{1}{6} \times \boxed{\frac{1}{6}}$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(\text{⚃}, \text{⚅} \mid \text{Fair}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \approx 0.028$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\boxed{\frac{4}{21}}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\boxed{\frac{6}{21}}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(⚃, ⚅ \mid \text{Fair}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \approx 0.028$

$$\Pr(⚃, ⚅ \mid \text{Biased}) = \frac{4}{21} \times \frac{6}{21} = \frac{24}{441} \approx 0.054$$

Bayesian Inference

Example: The biased-die problem

There is the possibility that the die are biased in a specific way:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\boxed{\frac{4}{21}}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\boxed{\frac{6}{21}}$

We generate some observations from a randomly selected die:

- 2 rolls with 4 and 6 pips

Probability of the observations: $\Pr(⚃, ⚅ \mid \text{Fair}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \approx 0.028$

$$\Pr(⚃, ⚅ \mid \text{Biased}) = \frac{4}{21} \times \frac{6}{21} = \frac{24}{441} \approx 0.054$$

So, the observations are ~2 times more likely under the biased hypothesis

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid ⚃, ⚅)}^{\text{posterior probability}} = \frac{\Pr(⚃, ⚅ \mid \text{Biased}) \times \Pr(\text{Biased})}{\Pr(⚃, ⚅)}$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \Pr(\text{Biased})}{\Pr(\text{⚃}, \text{⚅})}$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\Pr(\text{⚃}, \text{⚅})}$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\Pr(\text{⚃}, \text{⚅})}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$
$$\Pr(\text{Biased}) = 0.1$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\underbrace{\Pr(\text{⚃}, \text{⚅})}_{\text{marginal likelihood}}}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$
$$\Pr(\text{Biased}) = 0.1$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased}) \times \Pr(\underbrace{\text{Biased}) + \Pr(\text{⚃}, \text{⚅}}_{\text{marginal likelihood}} \mid \text{Fair}) \times \Pr(\text{Fair})}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$
$$\Pr(\text{Biased}) = 0.1$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased}) \times \underbrace{\Pr(\text{Biased}) + \Pr(\text{⚃}, \text{⚅}}_{\text{marginal likelihood}} \mid \text{Fair}) \times \Pr(\text{Fair})}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$

$$\Pr(\text{Biased}) = 0.1$$

likelihoods:

$$\Pr(\text{⚃}, \text{⚅} \mid \text{Fair}) = \frac{1}{36}$$

$$\Pr(\text{⚃}, \text{⚅} \mid \text{Biased}) = \frac{24}{441}$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid \text{⚃}, \text{⚅})}^{\text{posterior probability}} = \frac{\overbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\underbrace{\Pr(\text{⚃}, \text{⚅} \mid \text{Biased}) \times \Pr(\text{Biased}) + \Pr(\text{⚃}, \text{⚅} \mid \text{Fair}) \times \Pr(\text{Fair})}_{\text{marginal likelihood}}}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$

$$\Pr(\text{Biased}) = 0.1$$

likelihoods:

$$\Pr(\text{⚃}, \text{⚅} \mid \text{Fair}) = \frac{1}{36}$$

$$\Pr(\text{⚃}, \text{⚅} \mid \text{Biased}) = \frac{24}{441}$$

posterior probability:

$$\Pr(\text{Biased} \mid \text{⚃}, \text{⚅}) = \frac{\frac{24}{441} \times \frac{1}{10}}{\frac{24}{441} \times \frac{1}{10} + \frac{1}{36} \times \frac{9}{10}} \approx 0.18$$

Bayesian Inference

Example: The biased-die problem

What is the posterior probability of the alternative hypotheses?

$$\overbrace{\Pr(\text{Biased} \mid ⚃, ⚅)}^{\text{posterior probability}} = \frac{\overbrace{\Pr(⚃, ⚅ \mid \text{Biased})}^{\text{likelihood}} \times \overbrace{\Pr(\text{Biased})}^{\text{prior probability}}}{\Pr(⚃, ⚅ \mid \text{Biased}) \times \underbrace{\Pr(\text{Biased}) + \Pr(⚃,}_{\text{marginal likelihood}} ⚅ \mid \text{Fair}) \times \Pr(\text{Fair})}$$

prior probabilities:

$$\Pr(\text{Fair}) = 0.9$$

$$\Pr(\text{Biased}) = 0.1$$

likelihoods:

$$\Pr(⚃, ⚅ \mid \text{Fair}) = \frac{1}{36}$$

$$\Pr(⚃, ⚅ \mid \text{Biased}) = \frac{24}{441}$$

posterior probability:

$$\Pr(\text{Biased} \mid ⚃, ⚅) = \frac{\frac{24}{441} \times \frac{1}{10}}{\frac{24}{441} \times \frac{1}{10} + \frac{1}{36} \times \frac{9}{10}} \approx 0.18$$

So, our posterior belief in the biased hypothesis is $\Pr = 0.18$, which is an updated version of our prior belief in the biased hypothesis is $\Pr = 0.10$

Bayesian Inference

Generic statistical paradigm

pose a substantive question

develop a stochastic model with parameters that, if know, would answer the question

collect observations that are informative about model parameters

find the best estimate of model parameters (by some means) conditioned on (*i.e.*, given) the data at hand

Coin tossing

Is this a fair coin? Or, what is the probability of observing heads in a single toss?

Bayesian Inference

Generic statistical paradigm

pose a substantive question

develop a stochastic model with parameters that, if know, would answer the question

collect observations that are informative about model parameters

find the best estimate of model parameters (by some means) conditioned on (*i.e.*, given) the data at hand

Coin tossing

Is this a fair coin? Or, what is the probability of observing heads in a single toss?

Binomial probability distribution with parameter θ (probability of observing heads)

Bayesian Inference

Generic statistical paradigm	Coin tossing
pose a substantive question	Is this a fair coin? Or, what is the probability of observing heads in a single toss?
develop a stochastic model with parameters that, if know, would answer the question	Binomial probability distribution with parameter θ (probability of observing heads)
collect observations that are informative about model parameters	toss the coin n times and record the number of heads, x.
find the best estimate of model parameters (by some means) conditioned on (*i.e.*, given) the data at hand	

Bayesian Inference

Generic statistical paradigm	Coin tossing
pose a substantive question	Is this a fair coin? Or, what is the probability of observing heads in a single toss?
develop a stochastic model with parameters that, if know, would answer the question	Binomial probability distribution with parameter θ (probability of observing heads)
collect observations that are informative about model parameters	toss the coin n times and record the number of heads, x.
find the best estimate of model parameters (by some means) conditioned on (*i.e.*, given) the data at hand	find the best estimate of the θ parameter using Bayesian inference

Bayesian Inference

Example: Coin tossing

$$\underbrace{\boxed{\Pr(\theta \mid x)}}_{\text{posterior probability}} = \frac{\overbrace{\boxed{\Pr(x \mid \theta)}}^{\text{likelihood function}}\overbrace{\boxed{\Pr(\theta)}}^{\text{prior probability}}}{\underbrace{\boxed{\Pr(x)}}_{\text{marginal likelihood}}}$$

Bayesian Inference

Example: Coin tossing

likelihood function

prior probability

$$\boxed{\Pr(\theta \mid x)} = \frac{\boxed{\Pr(x \mid \theta)}\boxed{\Pr(\theta)}}{\boxed{\int \Pr(x \mid \theta)\Pr(\theta)d\theta}}$$

marginal likelihood

posterior probability

Bayesian Inference

Example: Coin tossing

likelihood function

prior probability

$$\boxed{\Pr(\theta \mid x)} = \frac{\boxed{\Pr(x \mid \theta)}\boxed{\Pr(\theta)}}{\boxed{\int \Pr(x \mid \theta)\Pr(\theta)\,d\theta}}$$

marginal likelihood

posterior probability

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}$$

heads

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

tails

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

possible orderings
of x heads in n tosses

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

possible orderings
of x heads in n tosses

This is called the Binomial coefficient, and is read 'n choose x':

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

The likelihood function for the Binomial distribution:

$$L(\theta; x) \propto \Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Bayesian Inference

Example: Coin tossing

We will adopt the Binomial distribution as our model of coin tossing:
discrete probability distribution that has two outcomes (*e.g.*, T/F, Y/N, H/T)

$$\Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

The likelihood function for the Binomial distribution:

$$L(\theta; x) \propto \Pr(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

With some algebra, we can solve for θ to find the MLE:

$$\hat{\theta} = \frac{x}{n}$$

Bayesian Inference

Example: Coin tossing

likelihood function

prior probability

$$\boxed{\Pr(\theta \mid x)} = \frac{\boxed{\Pr(x \mid \theta)}\boxed{\Pr(\theta)}}{\boxed{\int \Pr(x \mid \theta)\Pr(\theta)d\theta}}$$

marginal likelihood

posterior probability

Bayesian Inference

Example: Coin tossing

likelihood function

prior probability

$$\boxed{\Pr(\theta \mid x)} = \frac{\boxed{\Pr(x \mid \theta)}\boxed{\Pr(\theta)}}{\boxed{\int \Pr(x \mid \theta)\Pr(\theta)d\theta}}$$

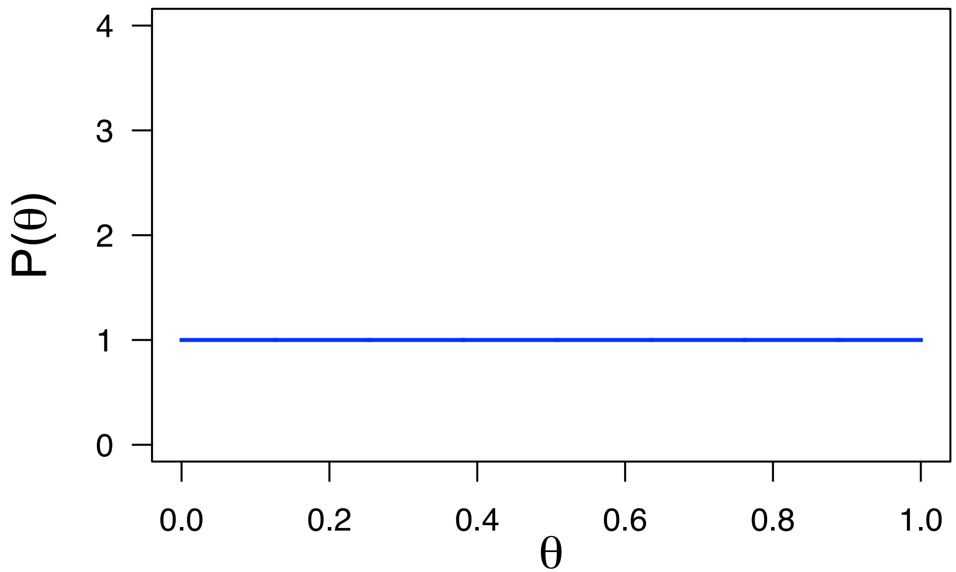
posterior probability

marginal likelihood

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:

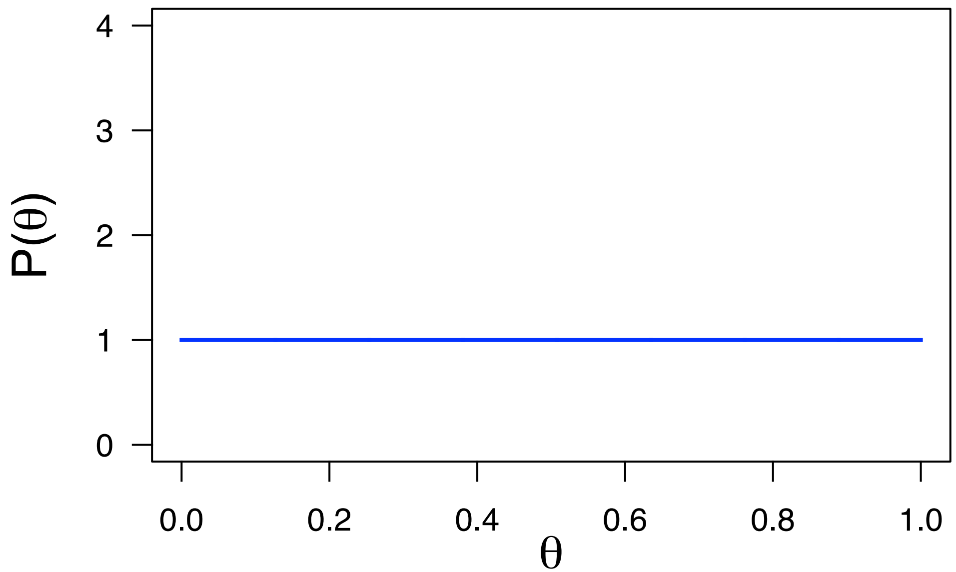


uniform prior: $\alpha = \beta = 1$

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:



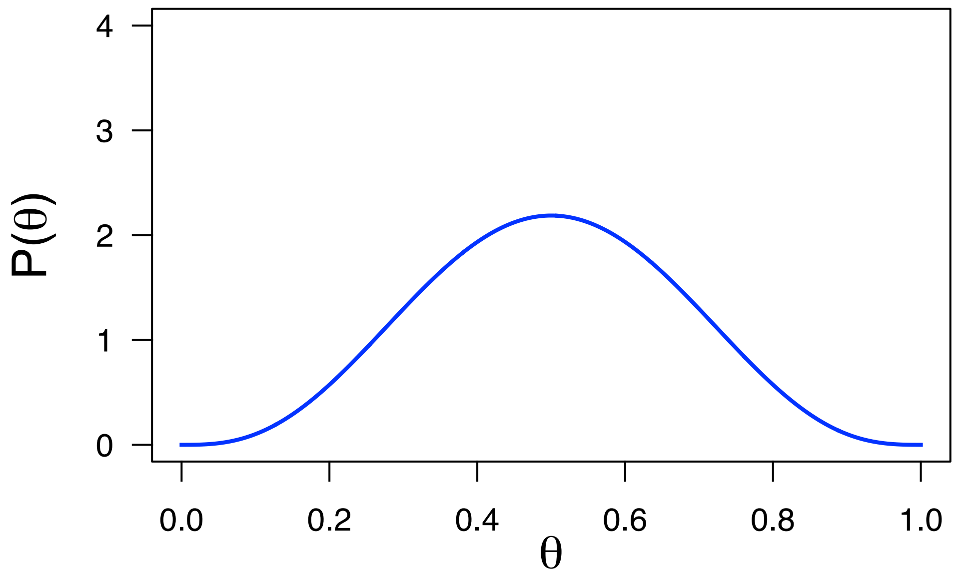
uniform prior: $\alpha = \beta = 1$

NOTE: uniform prior $\neq$ uninformative

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:

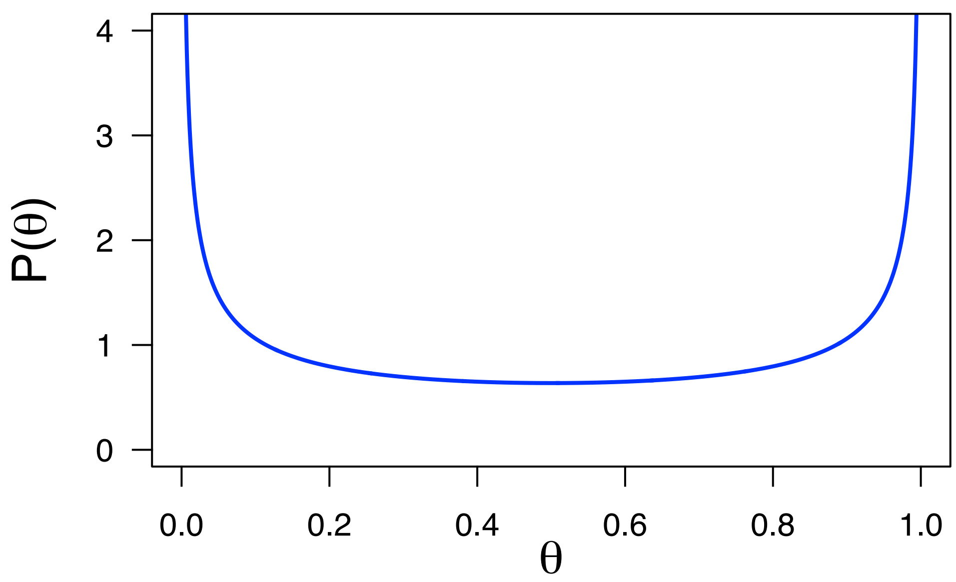


nonuniform prior: $\alpha = \beta = 4$

Bayesian Inference

Example: Coin tossing

The Beta prior probability distribution:

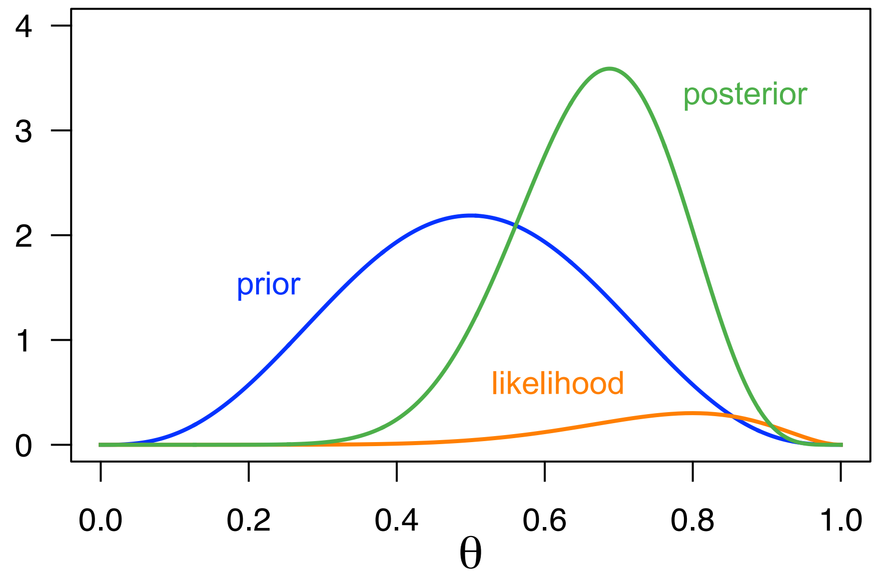


nonuniform prior: $\alpha = \beta = 0.5$

Bayesian Inference

Example: Coin tossing

The impact of the prior probability distribution on the estimated posterior probability:



$x = 8$ heads in $n = 10$ tosses

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Bayesian Inference of Phylogeny (on one slide)

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\quad \tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\quad \nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\quad \theta = \{a, b, c, d, e, f\}$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

$\nu_i \sim \text{Exponential}(\lambda = 10)$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

$\nu_i \sim \text{Exponential}(\lambda = 10)$

$\theta \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1)$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

$\nu_i \sim \text{Exponential}(\lambda = 10)$

$\theta \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1)$

$\pi \sim \text{Dirichlet}(1, 1, 1, 1)$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\quad \tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\quad \nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\quad \theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\quad \pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

$\nu_i \sim \text{Exponential}(\lambda = 10)$

$\theta \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1)$

$\pi \sim \text{Dirichlet}(1, 1, 1, 1)$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

V. Posterior Probability

$$\Pr(\tau, \nu, \theta, \pi \mid X) = \frac{\Pr(X \mid \tau, \nu, \theta, \pi) \Pr(\tau) \Pr(\nu) \Pr(\theta) \Pr(\pi)}{\Pr(X)}$$

Bayesian Inference of Phylogeny (on one slide)

I. Data

Assume an alignment, $\mathbf{X}$, of N sites for S species: $\mathbf{X} = (x_1, x_2, x_3, \ldots, x_N)$

II. Phylogenetic model parameters

1. Tree topology $\tau = \{\tau_1, \tau_2, \ldots, \tau_{(2S-5)!!}\}$

   branch lengths $\nu = \{\nu_1, \nu_2, \ldots, \nu_{2S-3}\}$

2. Model of character change

   relative substitution rates $\theta = \{a, b, c, d, e, f\}$

   stationary frequencies $\pi = \{\pi_A, \pi_C, \pi_G, \pi_T\}$

IV. Priors on parameters

$\tau \sim \text{Uniform}$

$\nu_i \sim \text{Exponential}(\lambda = 10)$

$\theta \sim \text{Dirichlet}(1,1,1,1,1,1)$

$\pi \sim \text{Dirichlet}(1,1,1,1)$

III. Phylogenetic likelihood function

$$\Pr(X \mid \tau, \nu, \theta, \pi) = \prod_{i=1}^{N} \Pr(x_i \mid \tau, \nu, \theta, \pi)$$

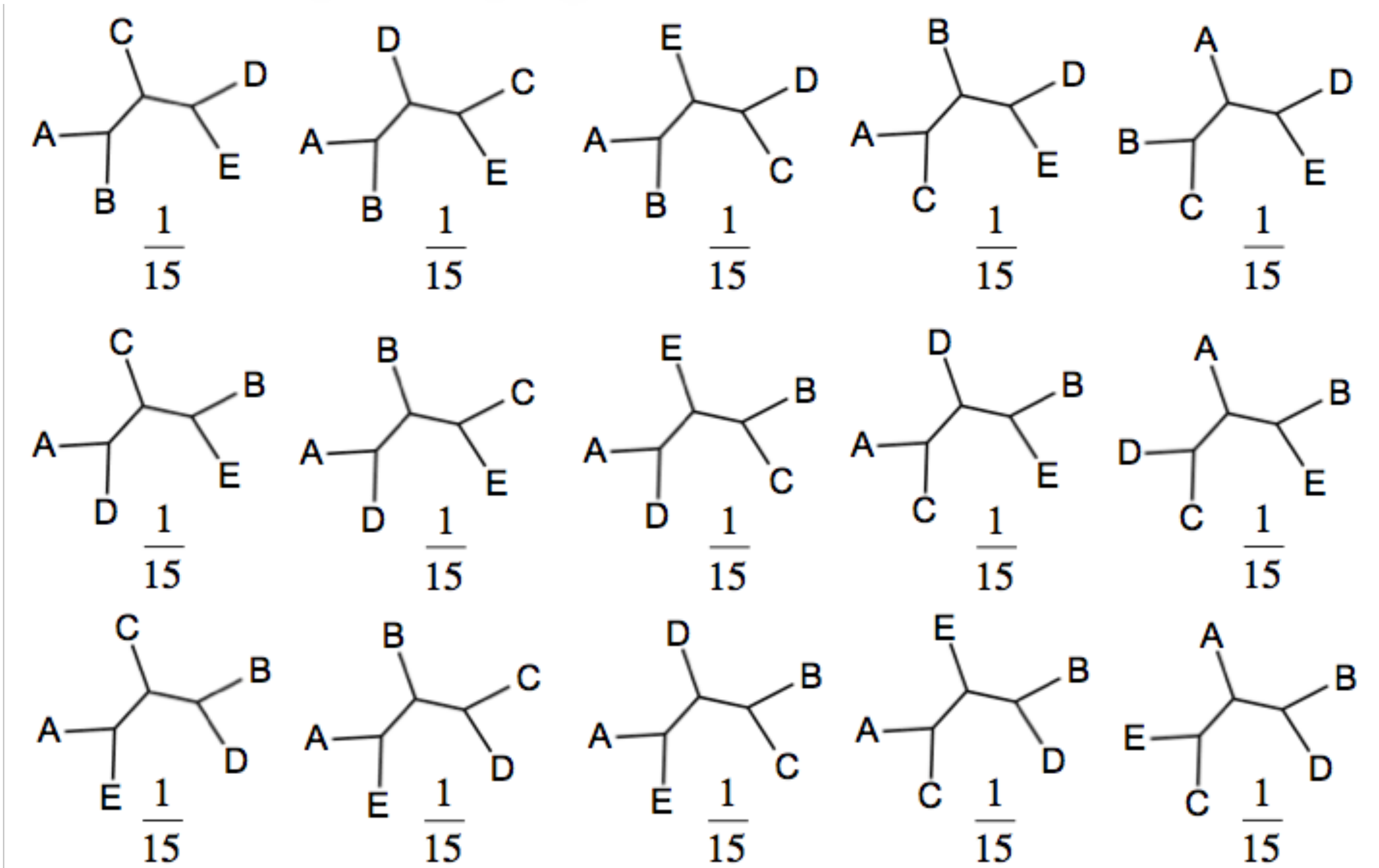
$$Q = \mu \begin{pmatrix} - & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & - & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & - & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & - \end{pmatrix}$$

V. Posterior Probability

$$\Pr(\tau, \nu, \theta, \pi \mid X) = \frac{\Pr(X \mid \tau, \nu, \theta, \pi)\ \Pr(\tau)\ \Pr(\nu)\ \Pr(\theta)\Pr(\pi)}{\Pr(X)}$$

Bayesian Inference of Phylogeny

Discrete-uniform prior on topologies



Bayesian Inference of Phylogeny

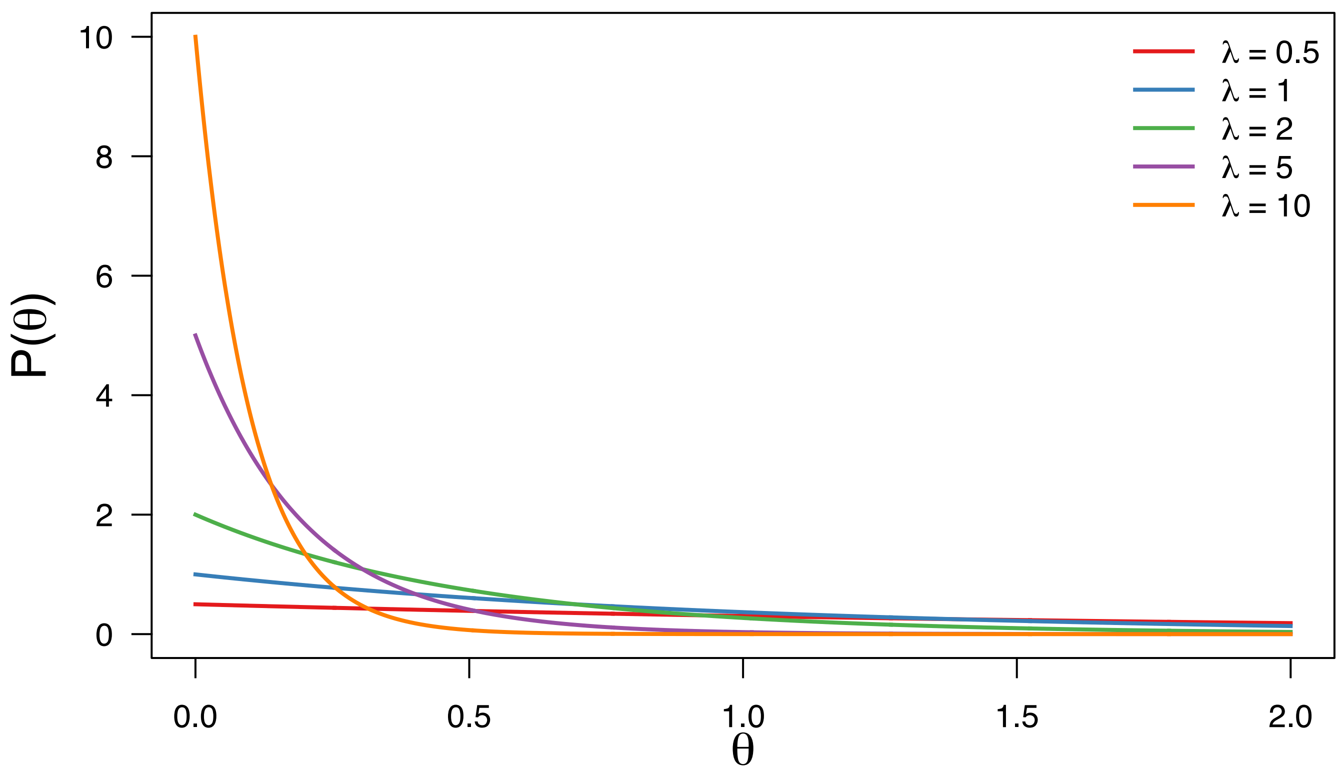
Exponential priors

Often used for branch lengths with rate parameter λ and mean $1/\lambda$

Bayesian Inference of Phylogeny

Exponential priors

Often used for branch lengths with rate parameter λ and mean $1/\lambda$



Bayesian Inference of Phylogeny

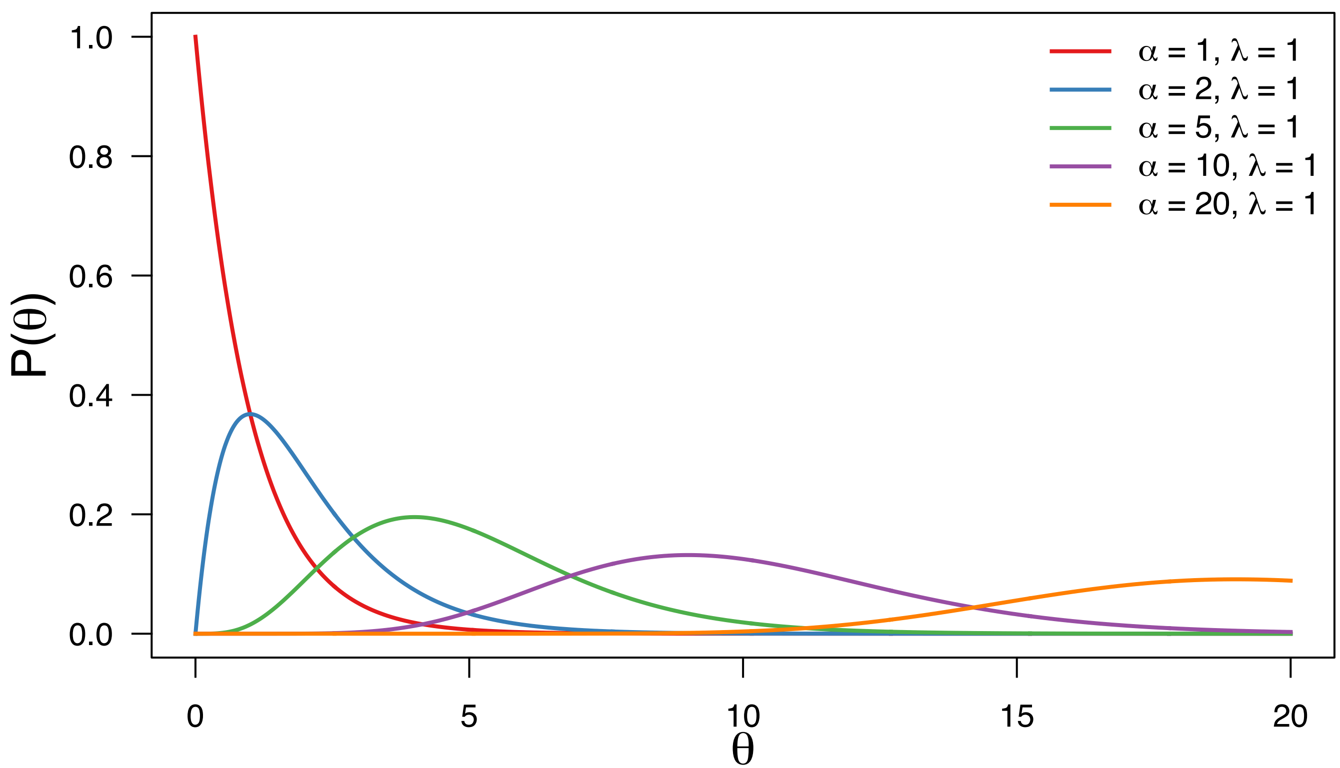
Gamma priors

A sum of α iid exponential variables

Bayesian Inference of Phylogeny

Gamma priors

A sum of α iid exponential variables



Bayesian Inference of Phylogeny

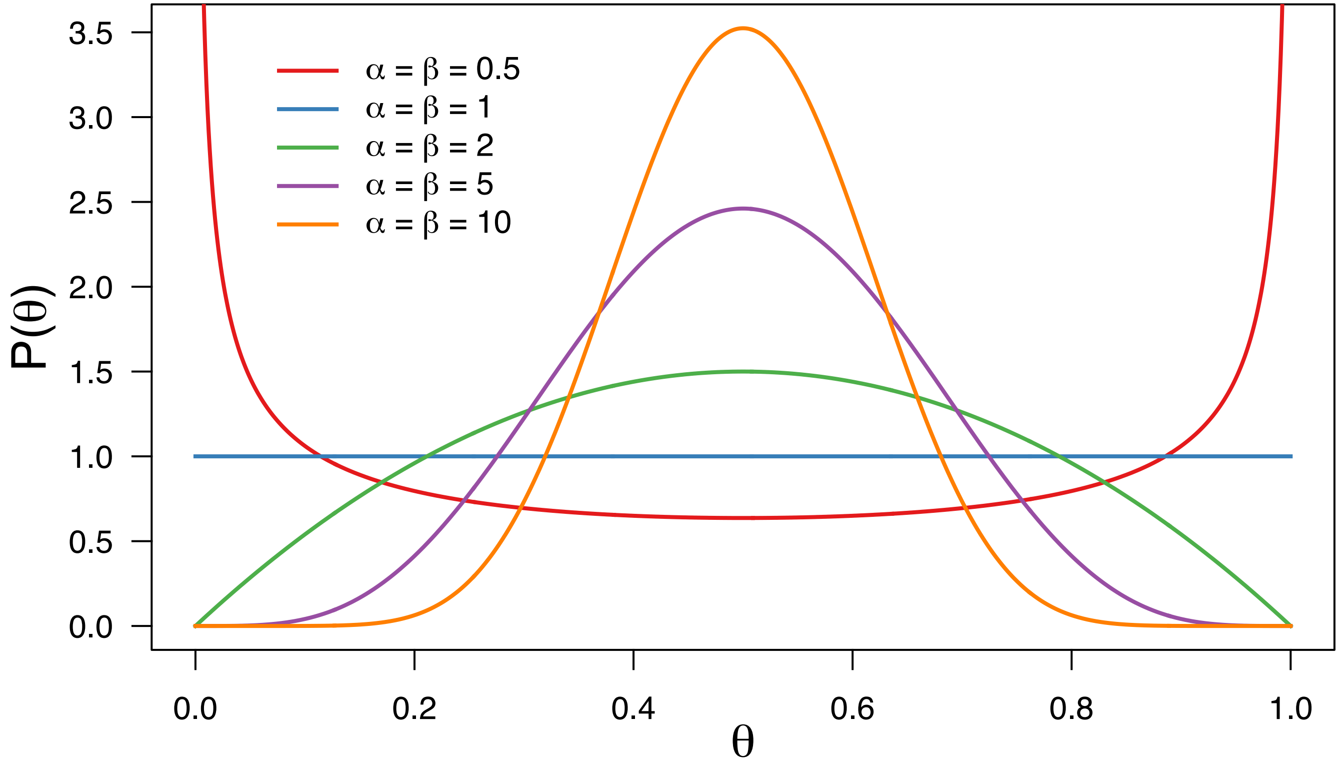
Beta prior

Often used for probabilities (like probability of heads) and fractions

Bayesian Inference of Phylogeny

Beta prior

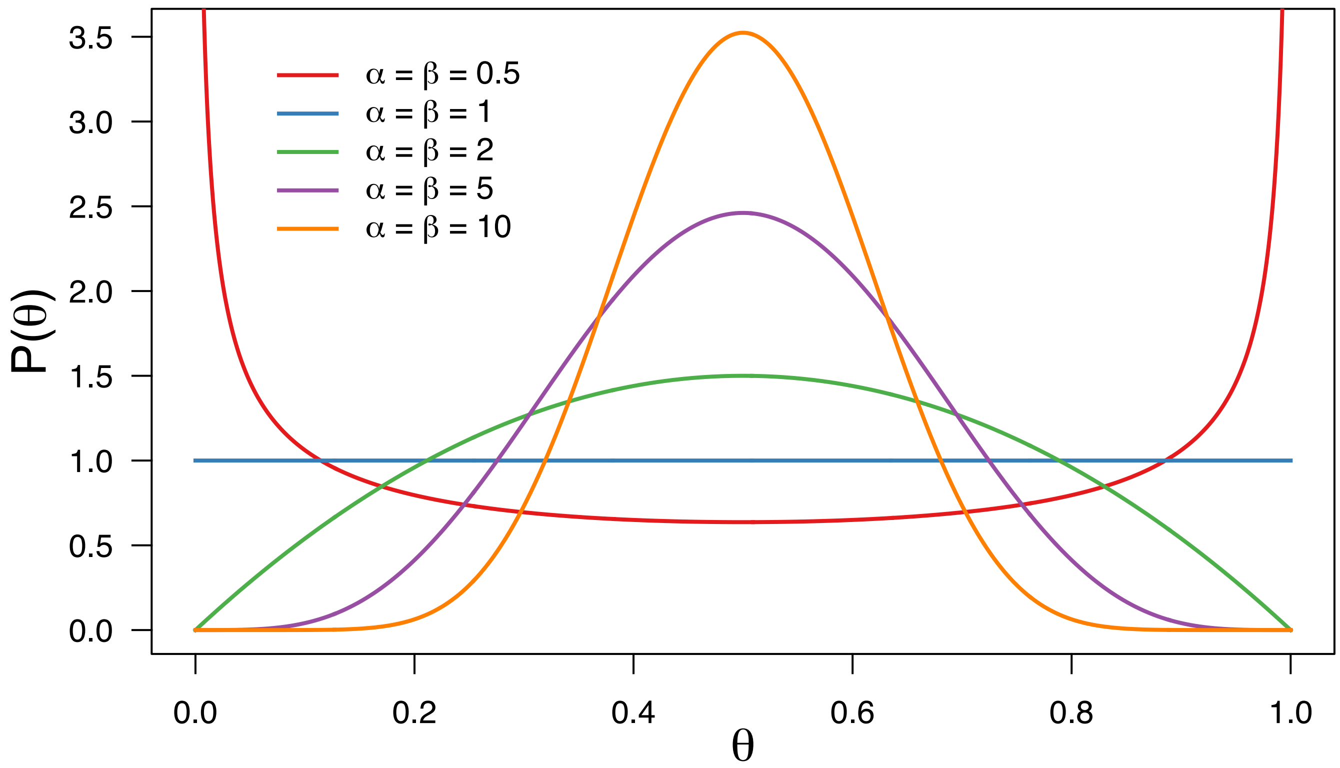
Often used for probabilities (like probability of heads) and fractions



Bayesian Inference of Phylogeny

Beta prior

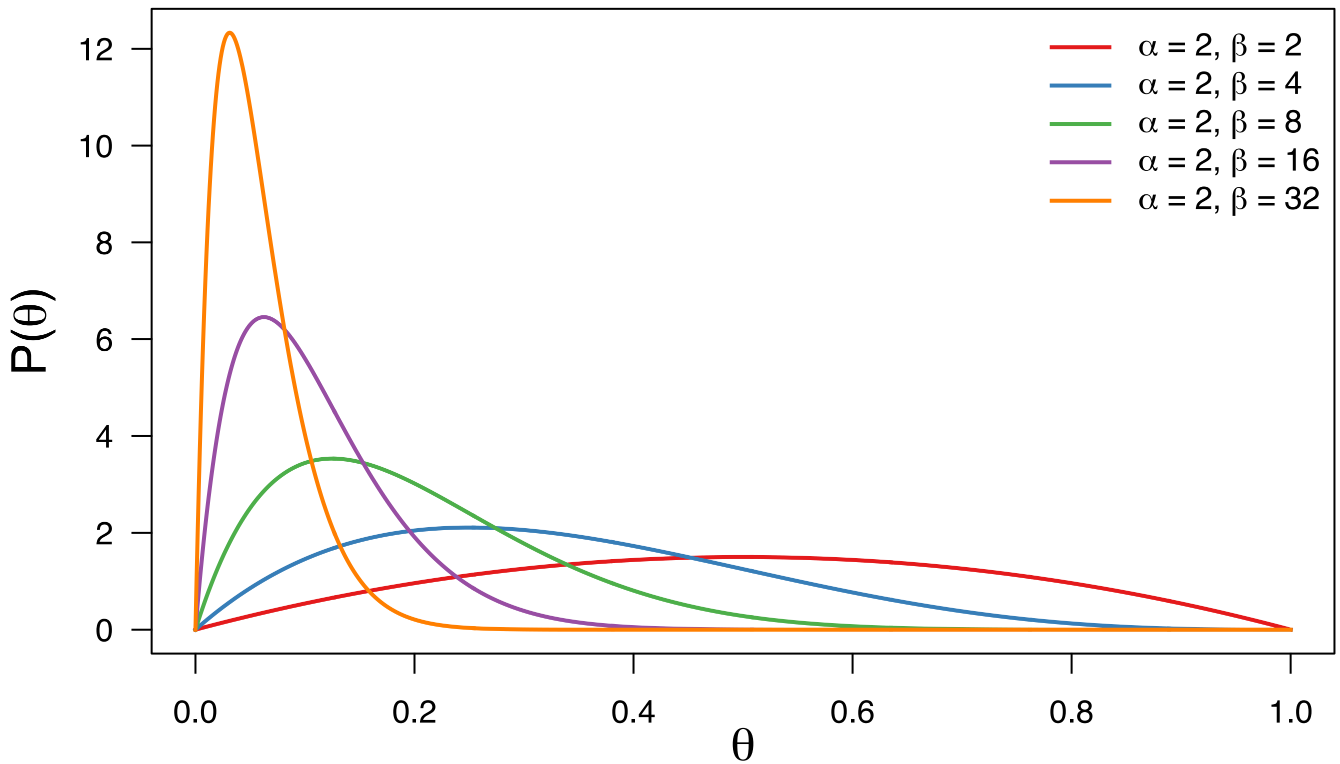
Often used for probabilities (like probability of heads) and fractions



Bayesian Inference of Phylogeny

Beta prior

Often used for probabilities (like probability of heads) and fractions



Bayesian Inference of Phylogeny

Dirichlet prior

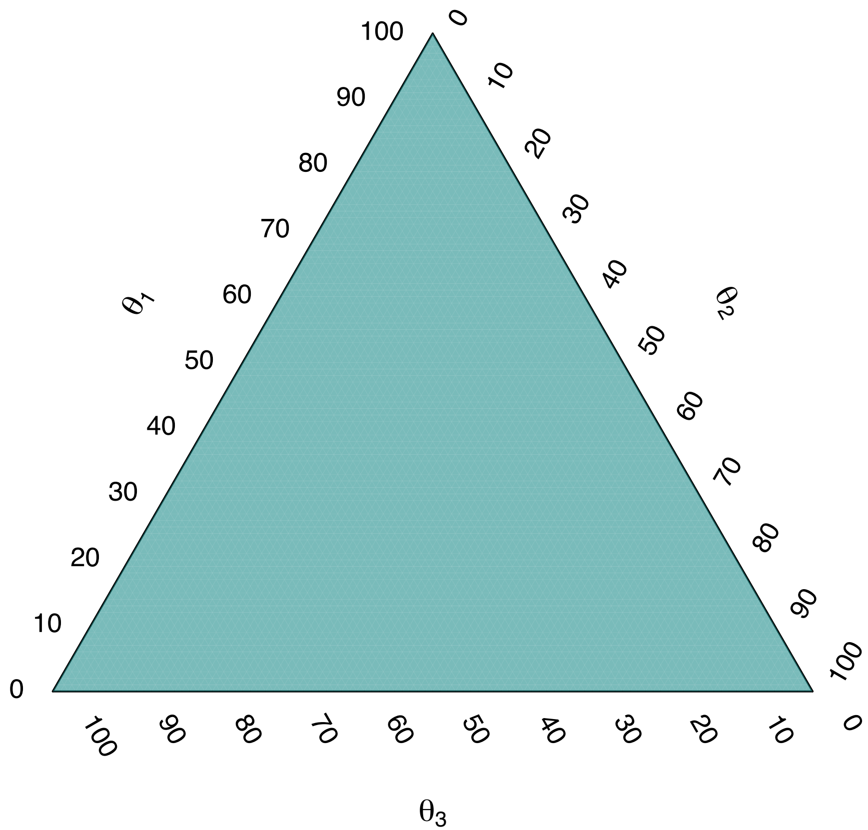
Generalization of the beta often used for proportions

Bayesian Inference of Phylogeny

Dirichlet prior

A “flat” Dirichlet distribution

$$\theta \sim \text{Dirichlet}(1, 1, 1)$$

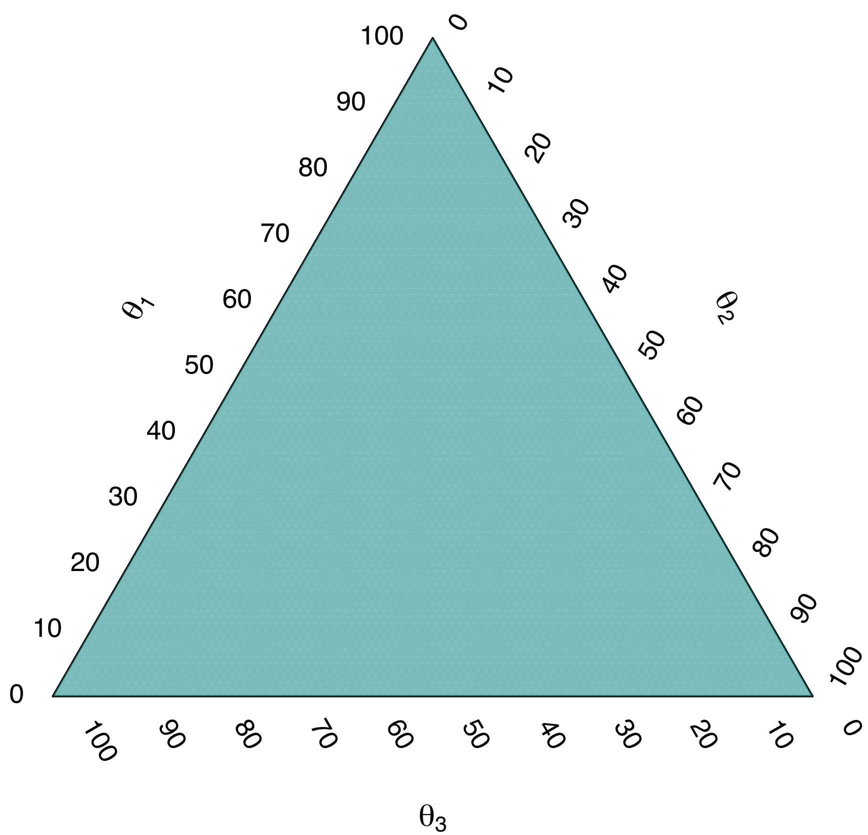


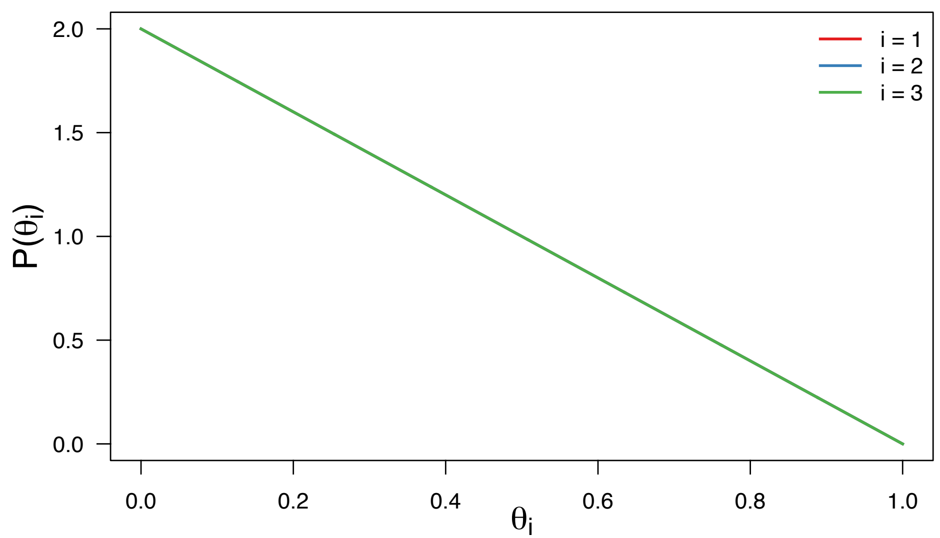
Bayesian Inference of Phylogeny

Dirichlet prior

A “flat” Dirichlet distribution

$$\theta \sim \mathrm{Dirichlet}(1,1,1)$$



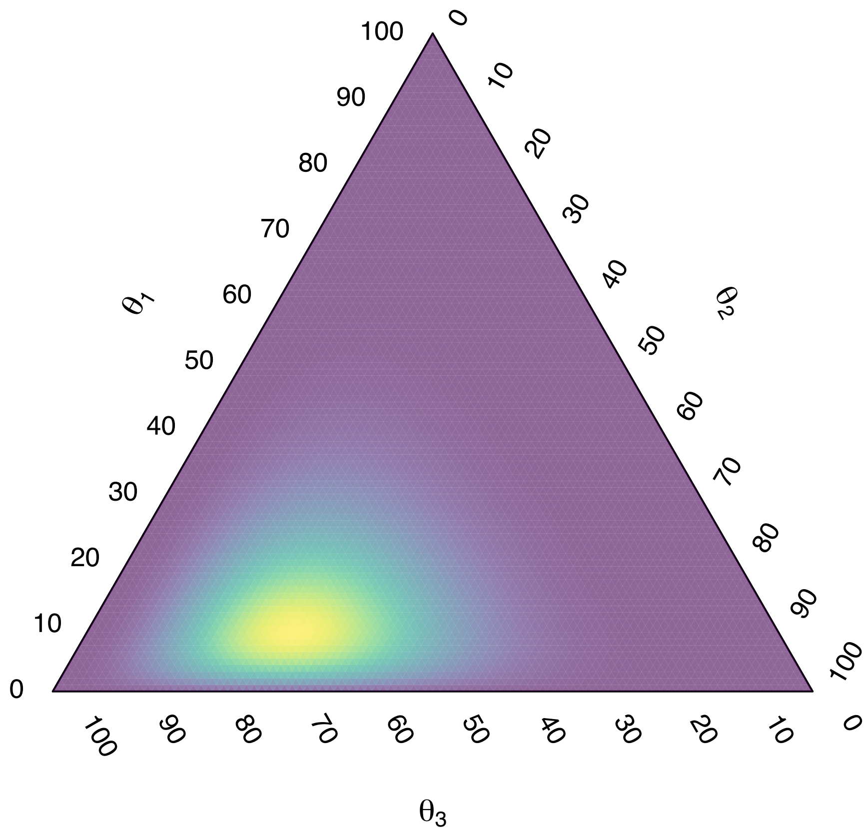


Bayesian Inference of Phylogeny

Dirichlet prior

An asymmetric Dirichlet distribution

$$\theta \sim \mathrm{Dirichlet}(2, 4, 8)$$

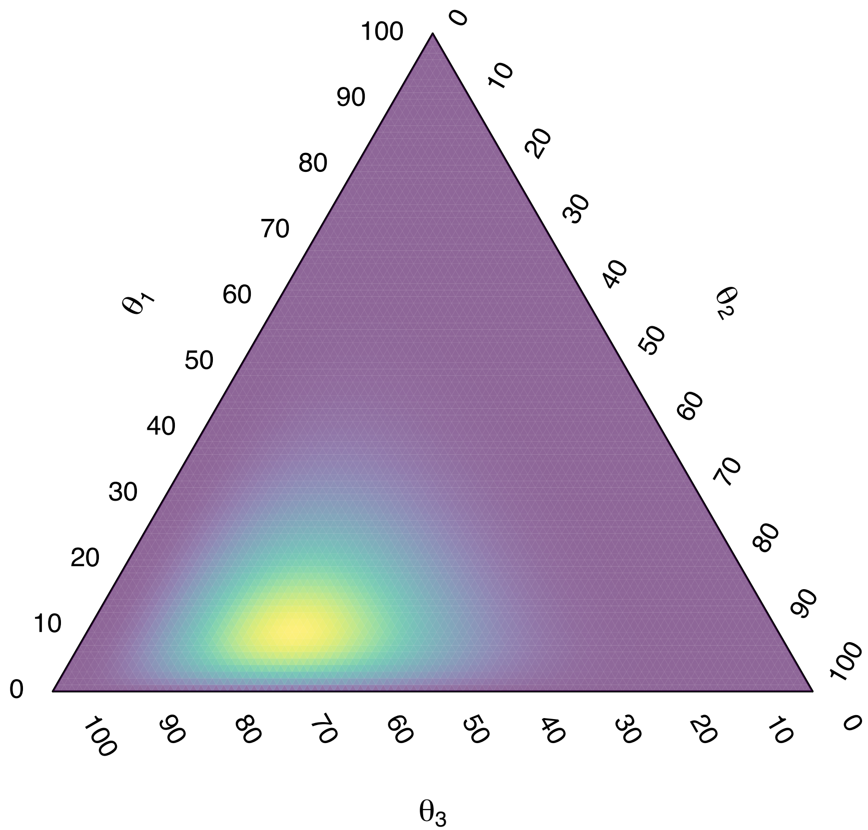


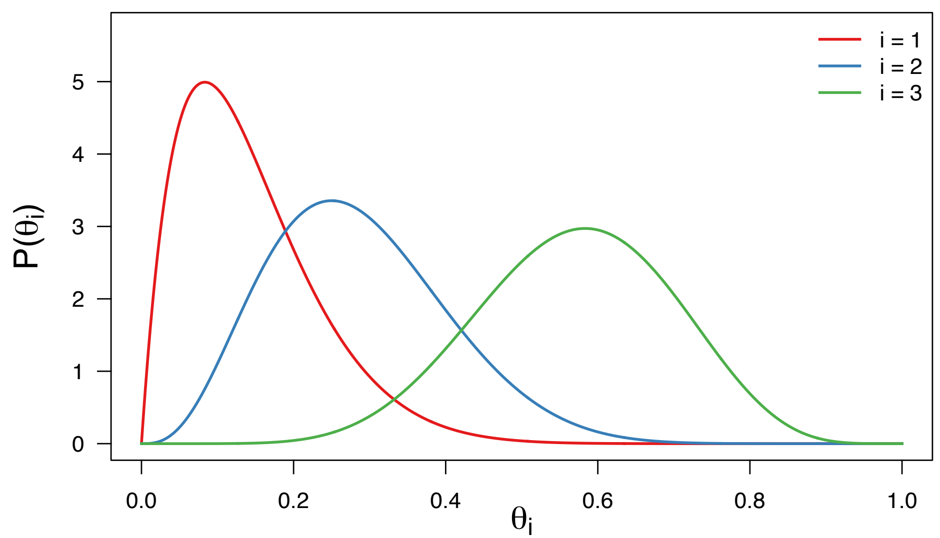
Bayesian Inference of Phylogeny

Dirichlet prior

An asymmetric Dirichlet distribution

$$\theta \sim \text{Dirichlet}(2, 4, 8)$$



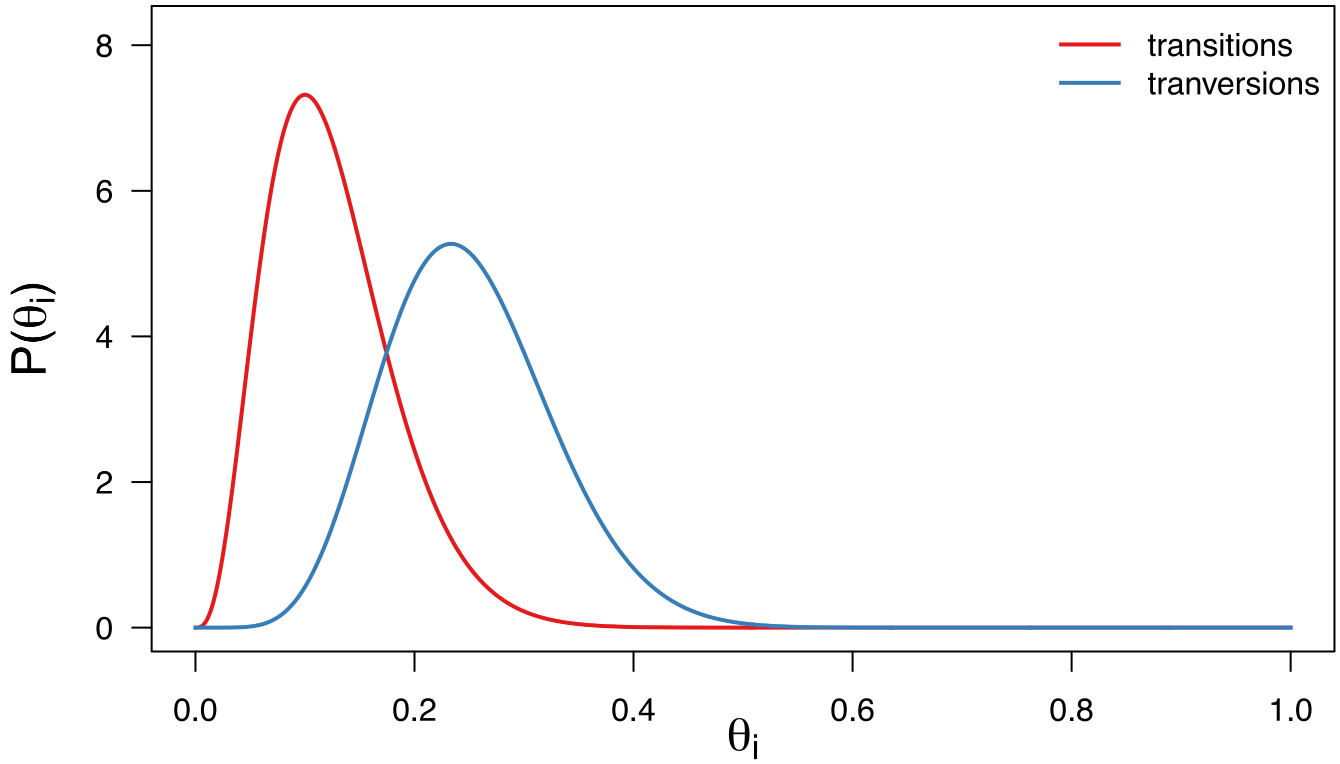


Bayesian Inference of Phylogeny

Dirichlet prior

We can express prior beliefs about transition/transversion rates

$$\theta \sim \text{Dirichlet}(4, 8, 4, 4, 8, 4)$$

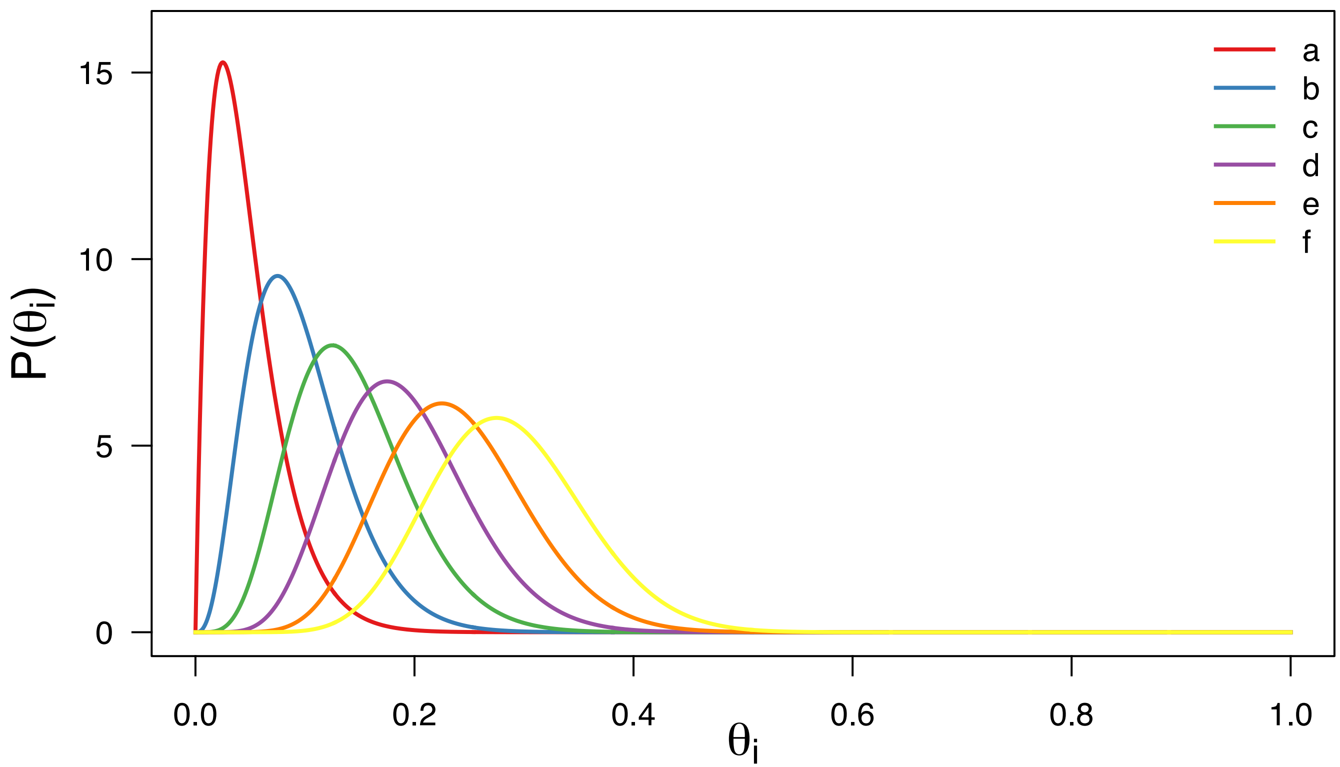


Bayesian Inference of Phylogeny

Dirichlet prior

Or any prior beliefs about exchangeability rates

$$\theta \sim \text{Dirichlet}(2, 4, 6, 8, 10, 12)$$



Bayesian Inference of Phylogeny

https://mikeryanmay.shinyapps.io/plotprior/

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Maximum Likelihood *vs.* Bayesian Inference

What is there to disagree about?

Not much, actually:

- model-based statistical inference

Maximum Likelihood *vs.* Bayesian Inference

What is there to disagree about?

Not much, actually:

- model-based statistical inference
- observations are random variables

Maximum Likelihood *vs.* Bayesian Inference

What is there to disagree about?

Not much, actually:

- model-based statistical inference
- observations are random variables
- likelihood function extracts information from data to estimate parameters

Maximum Likelihood *and* Bayesian Inference

What is there to disagree about?

Not much, actually:

- model-based statistical inference
- observations are random variables
- likelihood function extracts information from data to estimate parameters

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

- Desirable statistical properties

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Some less desirable aspects of maximum-likelihood estimation

Non-intuitive meaning of likelihood

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Some less desirable aspects of maximum-likelihood estimation

Non-intuitive meaning of likelihood

Frequentist perspective can be awkward for some inference problems

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Some less desirable aspects of maximum-likelihood estimation

Non-intuitive meaning of likelihood

Frequentist perspective can be awkward for some inference problems

Less amenable to EDA scenarios

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Some less desirable aspects of maximum-likelihood estimation

Non-intuitive meaning of likelihood

Frequentist perspective can be awkward for some inference problems

Less amenable to EDA scenarios

Accommodating uncertainty can be less than natural

Maximum Likelihood and Bayesian Inference

Lots to love about maximum-likelihood estimation

Desirable statistical properties

- consistent estimator
- asymptotically efficient estimator

Explicit with respect to model assumptions

Convenient and objective model selection/hypothesis testing framework

Some less desirable aspects of maximum-likelihood estimation

Non-intuitive meaning of likelihood

Frequentist perspective can be awkward for some inference problems

Less amenable to EDA scenarios

Accommodating uncertainty can be less than natural

Maximum Likelihood and Bayesian Inference

Maximum-likelihood perspective on parameters:

Data are random variables, but the parameters are fixed

Maximum Likelihood and Bayesian Inference

Maximum-likelihood perspective on parameters:

Data are random variables, but the parameters are fixed

Bayesian perspective on parameters:

Data are random variables, and so are the model parameters

Maximum Likelihood and Bayesian Inference

Maximum-likelihood perspective on parameters:

Data are random variables, but the parameters are fixed

Bayesian perspective on parameters:

Data are random variables, and so are the model parameters

If we treat the parameters as random variables, what do we have to specify?

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?

- this is consistent with making assumptions clear (model-based inference)

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?

- this is consistent with making assumptions clear (model-based inference)

When relevant prior information is available, it can be naturally incorporated

- this is consistent with the way we behave as rational beings

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?

- this is consistent with making assumptions clear (model-based inference)

When relevant prior information is available, it can be naturally incorporated

- this is consistent with the way we behave as rational beings

It can be non-trivial to specify our prior beliefs as probability distributions

- we might attempt to define vague priors in some cases

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?

- this is consistent with making assumptions clear (model-based inference)

When relevant prior information is available, it can be naturally incorporated

- this is consistent with the way we behave as rational beings

It can be non-trivial to specify our prior beliefs as probability distributions

- we might attempt to define vague priors in some cases
- we can (and should) assess the impact of our prior assumptions

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?
- this is consistent with making assumptions clear (model-based inference)

When relevant prior information is available, it can be naturally incorporated
- this is consistent with the way we behave as rational beings

It can be non-trivial to specify our prior beliefs as probability distributions
- we might attempt to define vague priors in some cases
- we can (and should) assess the impact of our prior assumptions

Concerns about the prior sensitivity are somewhat philosophical
- the posterior is typically dominated by the likelihood function

Bayesian Inference

A priori…

We usually (*i.e.*, always) have prior beliefs, so why not be explicit about it?

- this is consistent with making assumptions clear (model-based inference)

When relevant prior information is available, it can be naturally incorporated

- this is consistent with the way we behave as rational beings

It can be non-trivial to specify our prior beliefs as probability distributions

- we might attempt to define vague priors in some cases
- we can (and should) assess the impact of our prior assumptions

Concerns about the prior sensitivity are somewhat philosophical

- the posterior is typically dominated by the likelihood function
- when this is not the case, the ability to detect prior sensitivity is a good thing!

Bayesian Inference

is my prior ***informative***?

is my prior ***informed***?		no	yes
	no		
	yes		

Bayesian Inference

is my prior ***informative***?

is my prior ***informed***?		no	yes
	no	😀	
	yes		

Bayesian Inference

is my prior ***informative***?

is my prior ***informed***?

	no	yes
no	😀	
yes	😀	

Bayesian Inference

is my prior ***informative***?

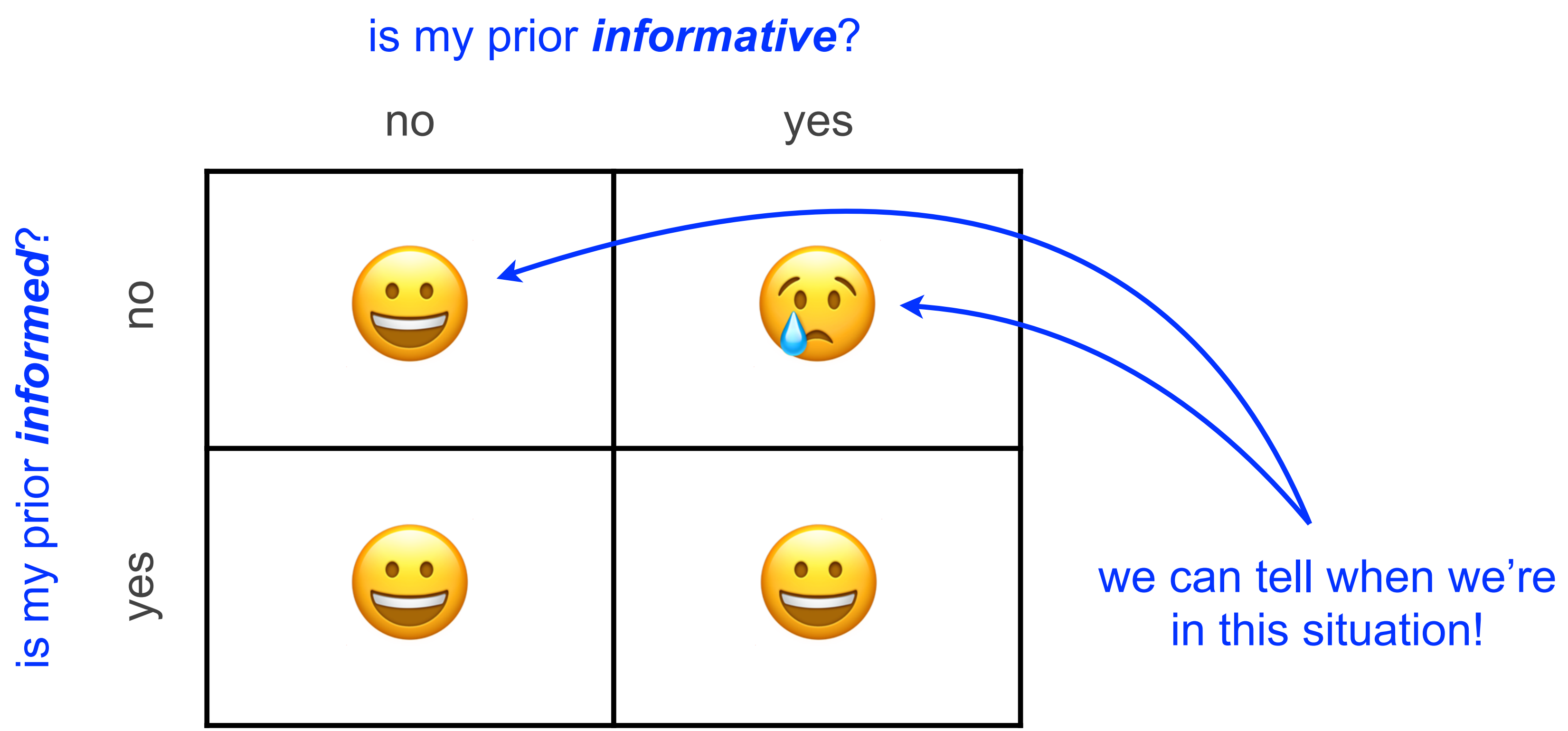
is my prior ***informed***?	no	yes
no	😀	
yes	😀	😀

Bayesian Inference

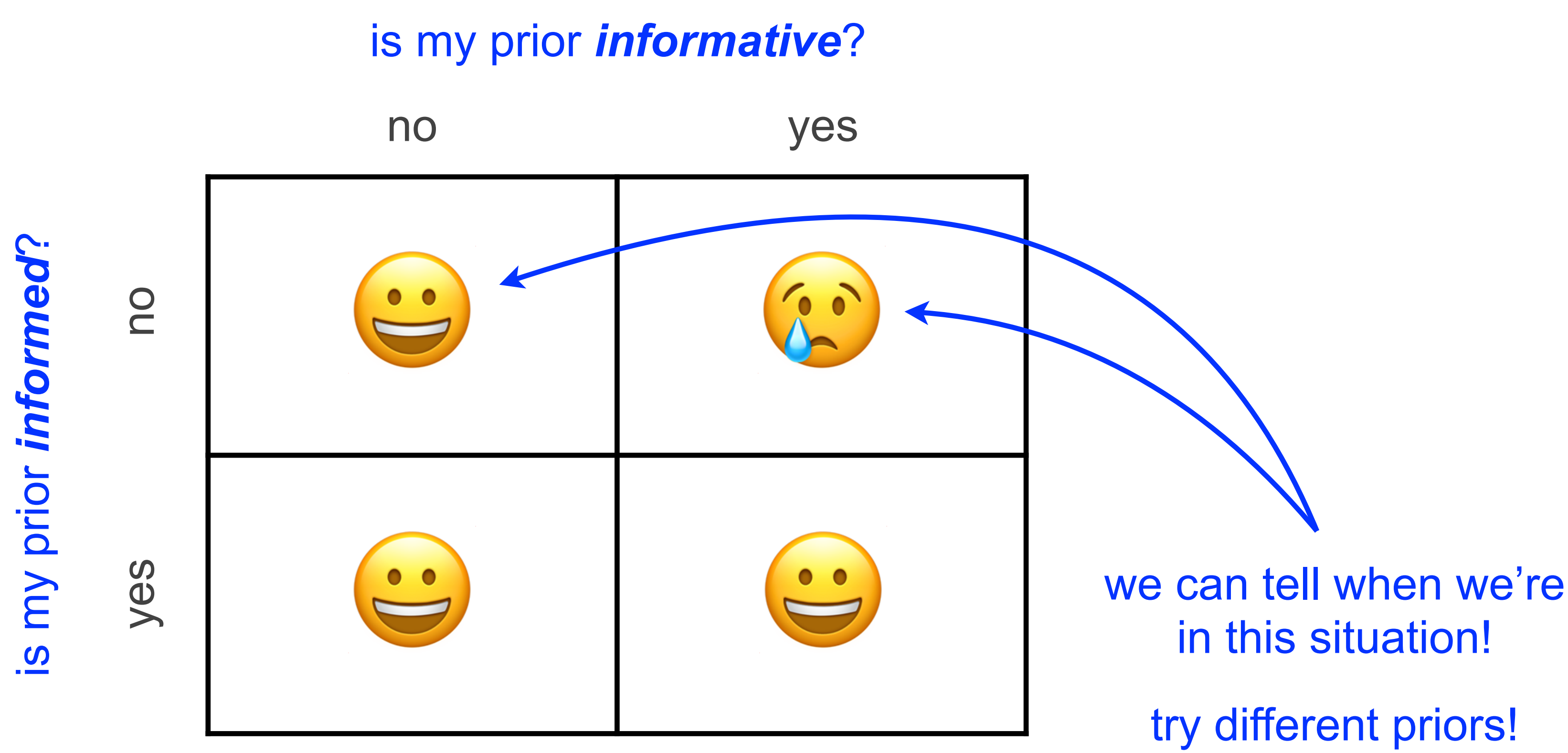
is my prior ***informative***?

is my prior ***informed***?		no	yes
	no	😀	😢
	yes	😀	😀

Bayesian Inference



Bayesian Inference



Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

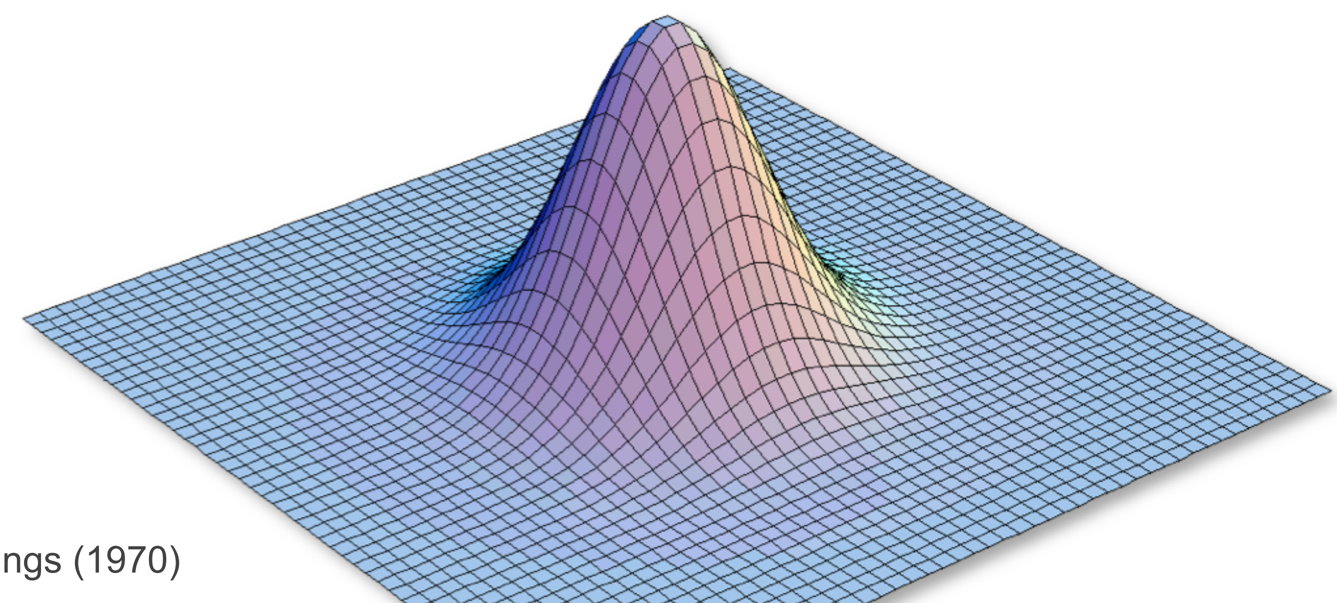
Approximating the Joint Posterior Probability Density using MCMC

Approximating the Joint Posterior Probability Density using 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:

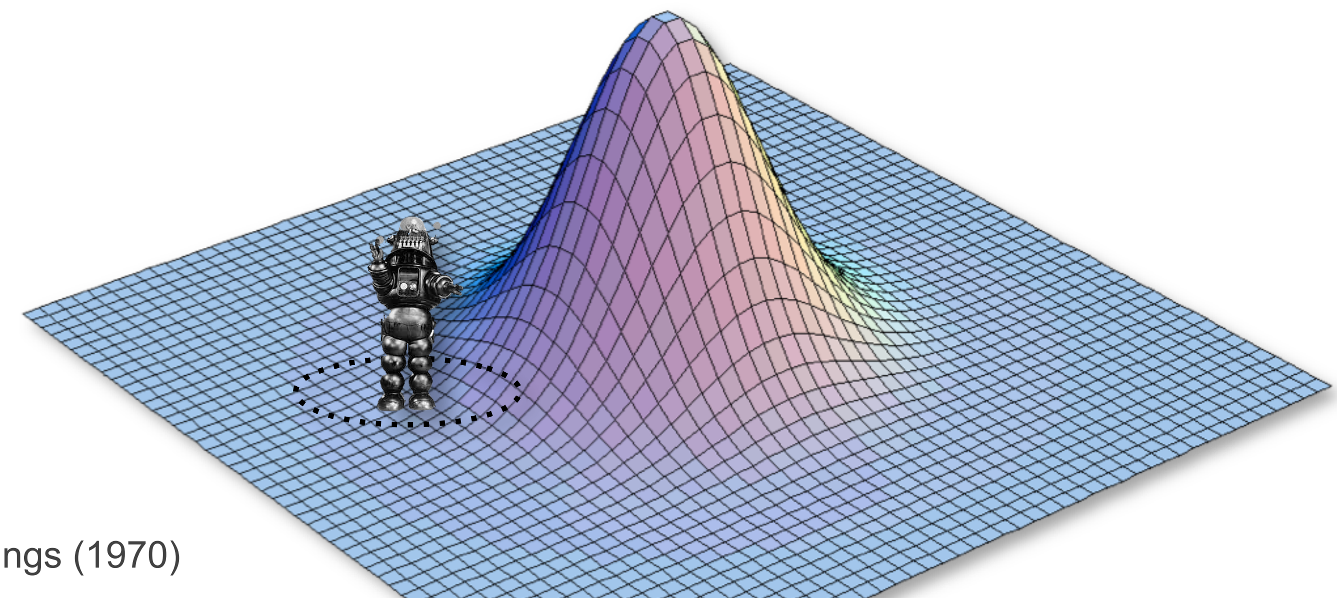
Metropolis et al. (1953); Hastings (1970)



Approximating the Joint Posterior Probability Density using 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)

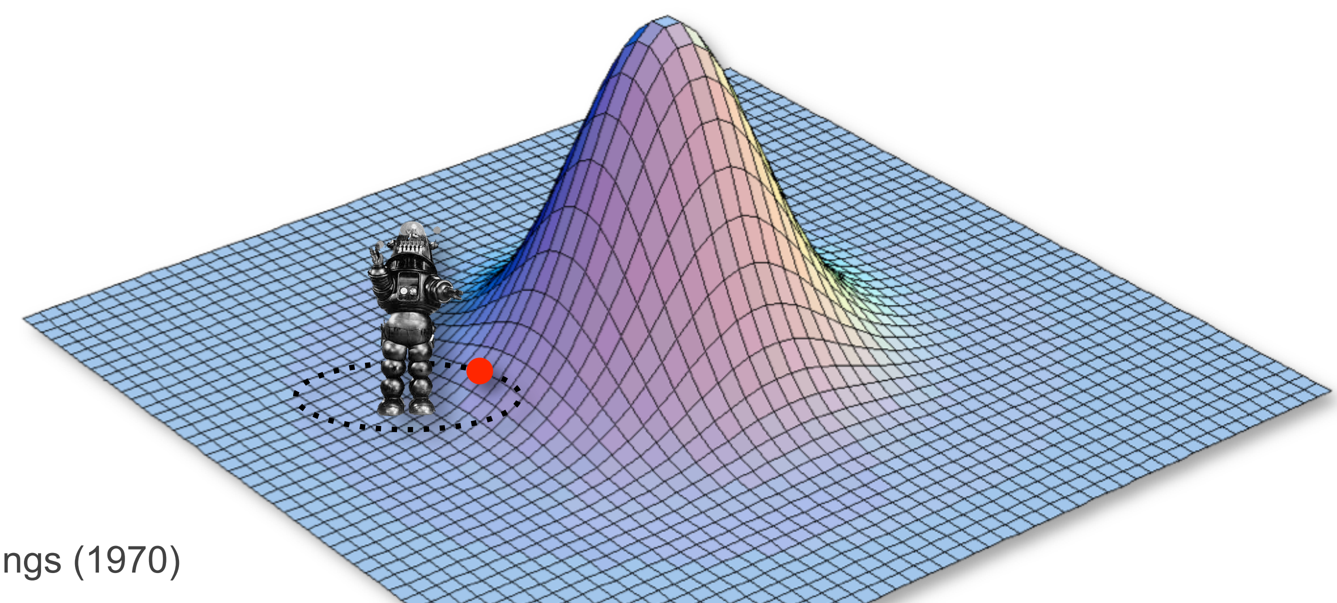
Approximating the Joint Posterior Probability Density using MCMC

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

$$\Pr(\text{Accept}) = 1$$



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

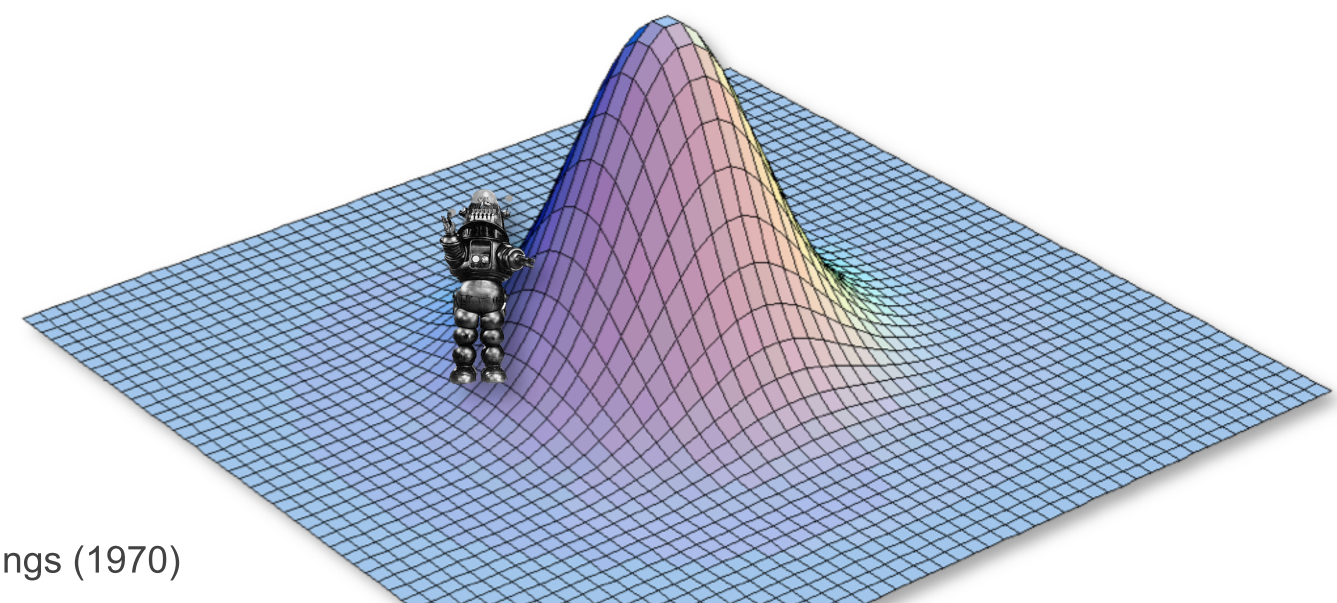
Approximating the Joint Posterior Probability Density using MCMC

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

$$\Pr(\text{Accept}) = 1$$



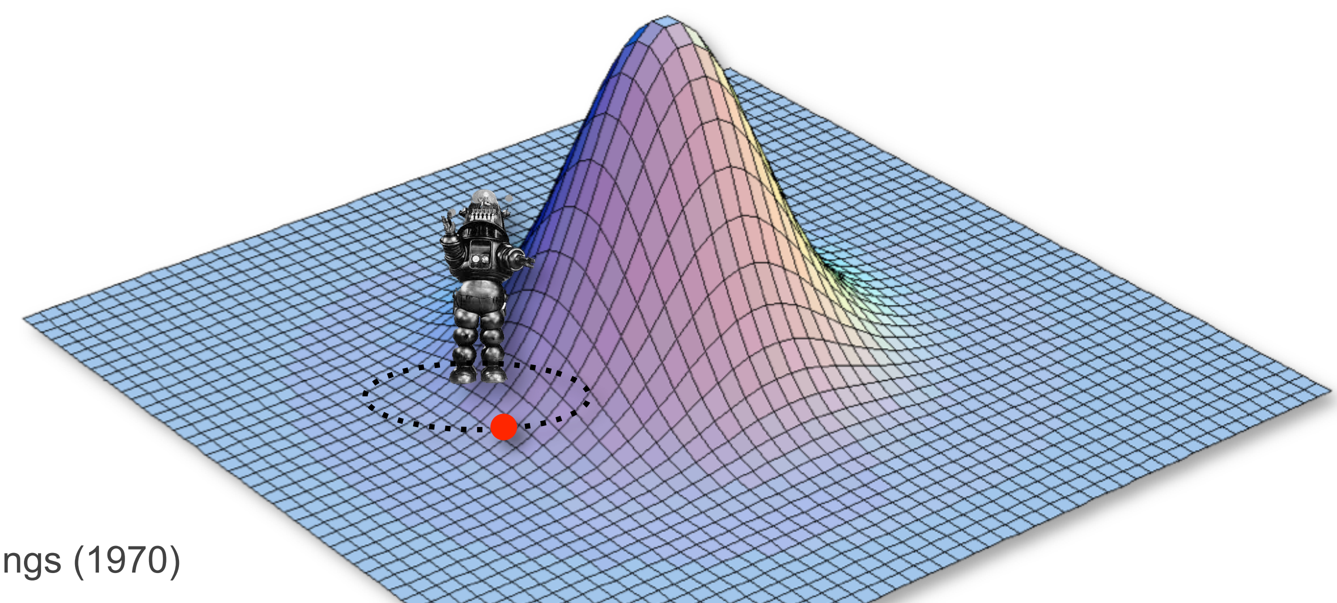
Metropolis et al. (1953); Hastings (1970)

Approximating the Joint Posterior Probability Density using 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, $u \sim \text{Uniform}(0,1)$



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

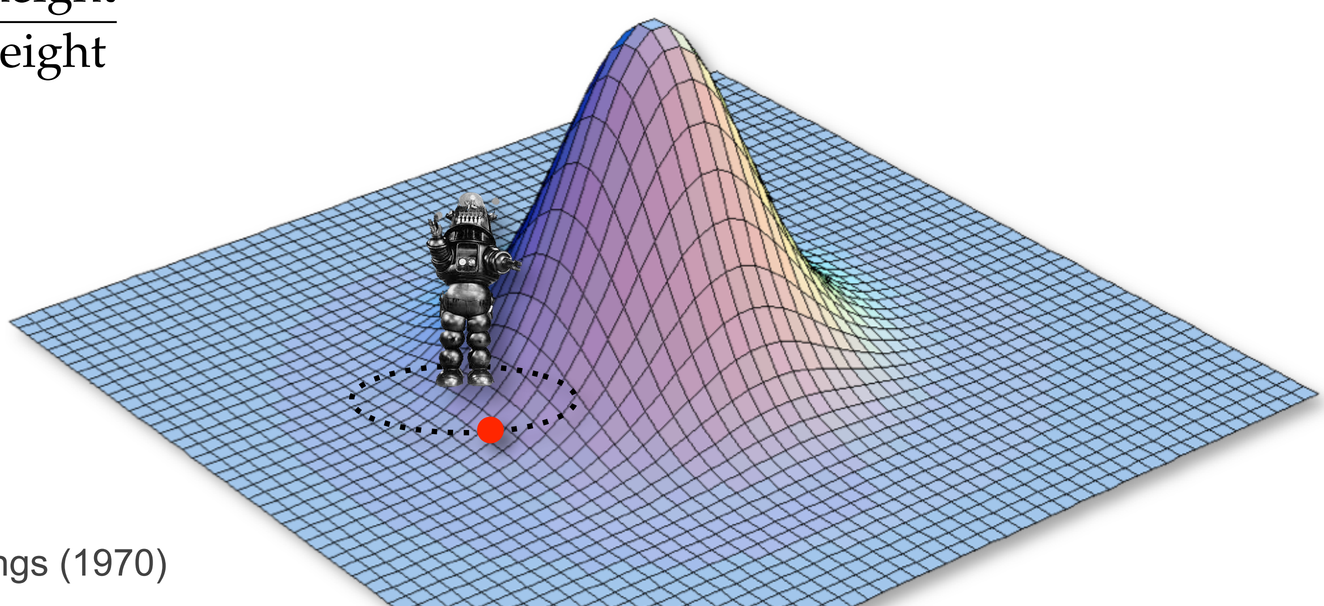
Approximating the Joint Posterior Probability Density using MCMC

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, $u \sim \text{Uniform}(0,1)$

$$\Pr(\text{Accept}) = \frac{\text{new height}}{\text{old height}}$$



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

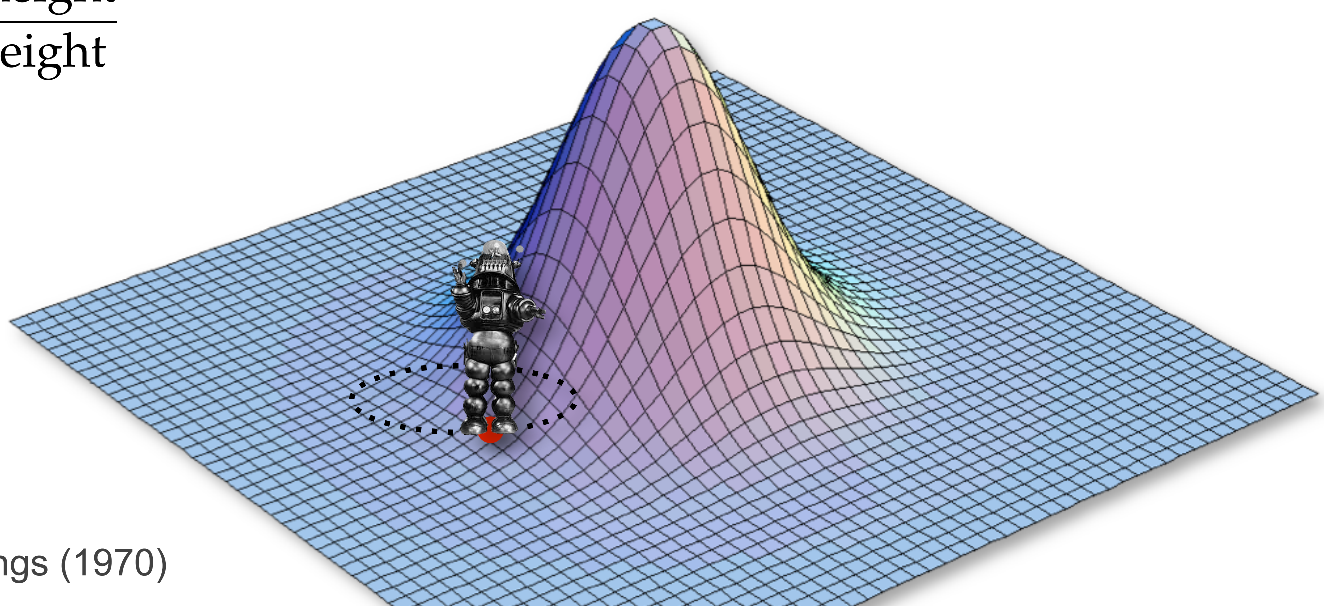
Approximating the Joint Posterior Probability Density using MCMC

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, $u \sim \text{Uniform}(0,1)$
3. Assume the proposal distribution is symmetrical, so $\text{Pr}(A \rightarrow B) = \text{Pr}(B \rightarrow A)$

$$\Pr(\text{Accept}) = \frac{\text{new height}}{\text{old height}}$$



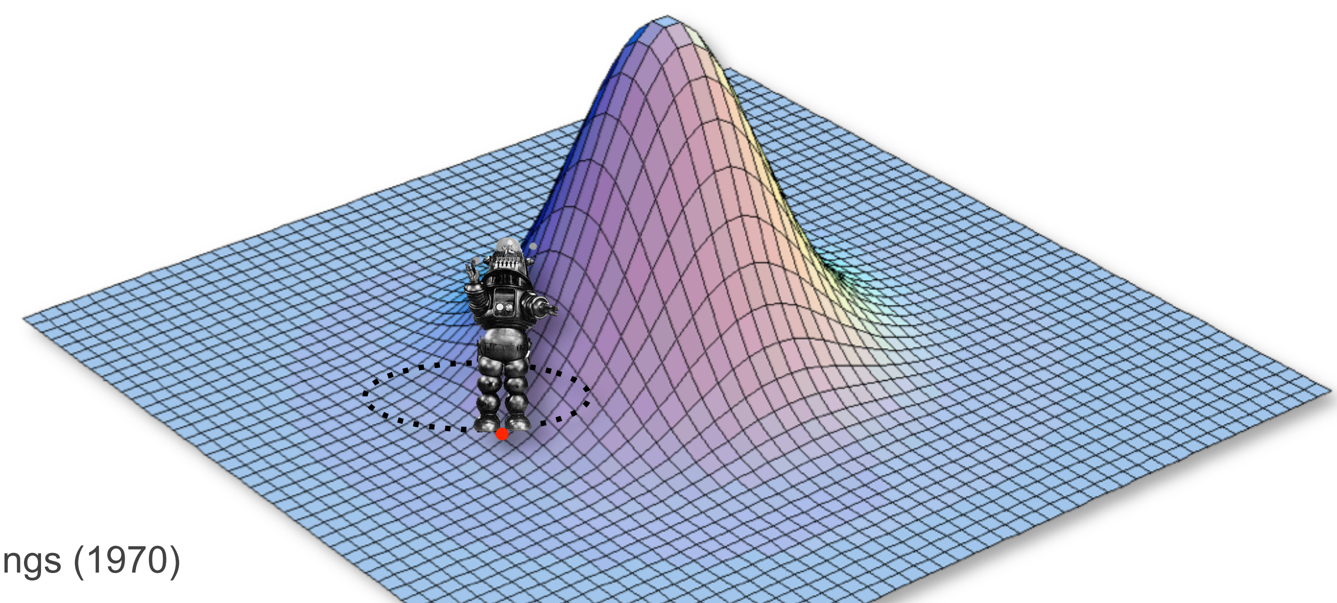
Metropolis et al. (1953); Hastings (1970)

Approximating the Joint Posterior Probability Density using 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, $u \sim \text{Uniform}(0,1)$
3. Assume the proposal distribution is symmetrical, so $\Pr(A \rightarrow B) = \Pr(B \rightarrow A)$



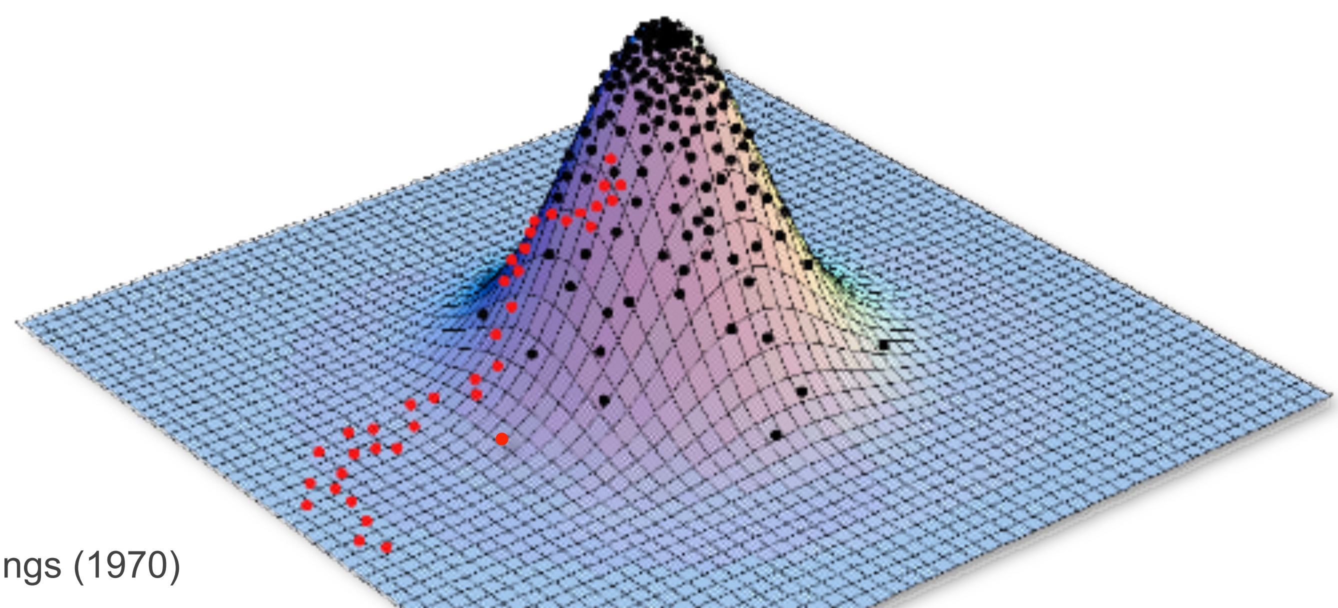
Metropolis et al. (1953); Hastings (1970)

Approximating the Joint Posterior Probability Density using 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, $u \sim$ Uniform$(0,1)$
3. Assume the proposal distribution is symmetrical, so Pr$(A \rightarrow B) =$ Pr$(B \rightarrow A)$



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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities

```
# specify a beta prior on x
x ~ dnBeta(1,1)

# place a sliding move on x
moves.append( mvSlide(x, delta = 0.1, weight = 5.0) )
```

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities

prior

parameter

```
# specify a beta prior on x
x ~ dnBeta(1,1)

# place a sliding move on x
moves.append( mvSlide(x, delta = 0.1, weight = 5.0) )
```

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities

prior

parameter

```
# specify a beta prior on x
x ~ dnBeta(1,1)

# place a sliding move on x
moves.append( mvSlide(x, delta = 0.1, weight = 5.0) )
```

proposal weight

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism

prior

parameter

```
# specify a beta prior on x
x ~ dnBeta(1,1)

# place a sliding move on x
moves.append( mvSlide(x, delta = 0.1, weight = 5.0) )
```

proposal weight

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change

$$R = \min\left[1, \underbrace{\boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)}}}_{\text{likelihood ratio}} \times \underbrace{\boxed{\frac{\Pr(\theta')}{\Pr(\theta)}}}_{\text{prior ratio}} \times \underbrace{\boxed{\frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}}}_{\text{proposal ratio}}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?

$$R = \min\left[1, \boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)}} \times \boxed{\frac{\Pr(\theta')}{\Pr(\theta)}} \times \boxed{\frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}}\right]$$

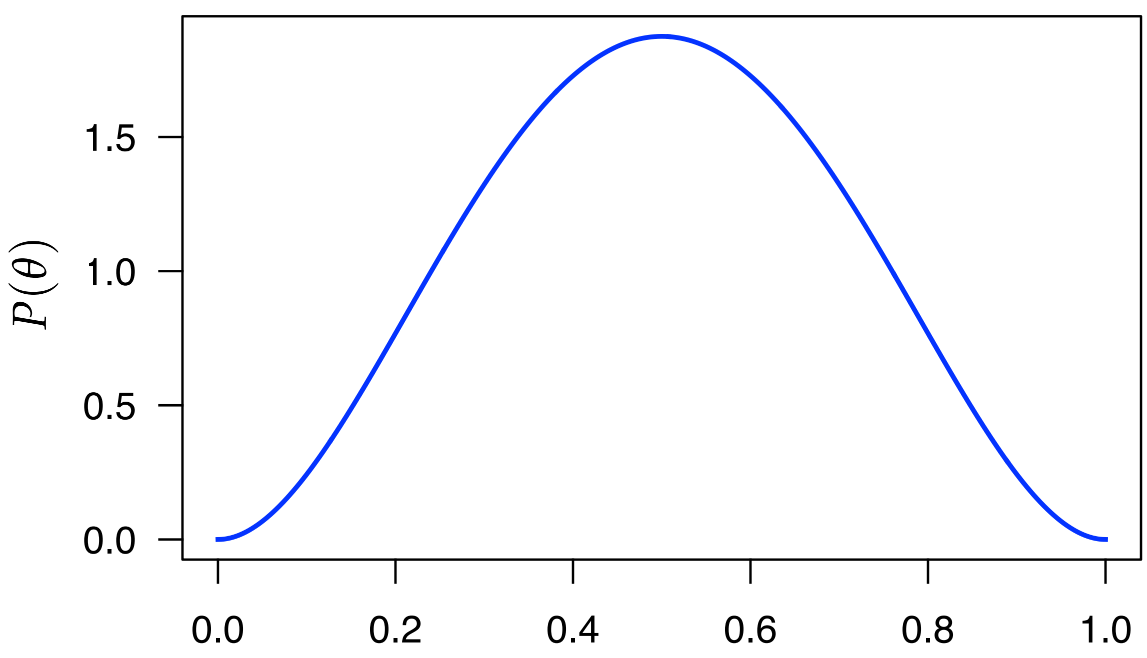
likelihood ratio — prior ratio — proposal ratio

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

Approximating the Joint Posterior Probability Density using MCMC

The prior for each parameter is specified

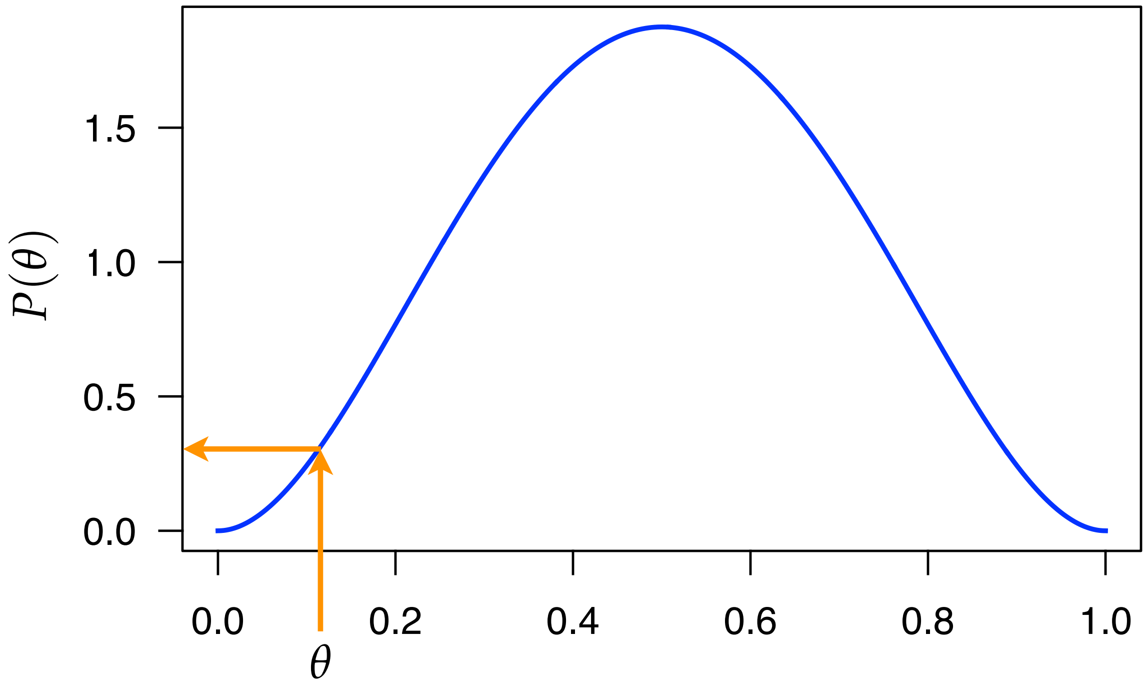
We can just look up the prior probability of a given parameter value



Approximating the Joint Posterior Probability Density using MCMC

The prior for each parameter is specified

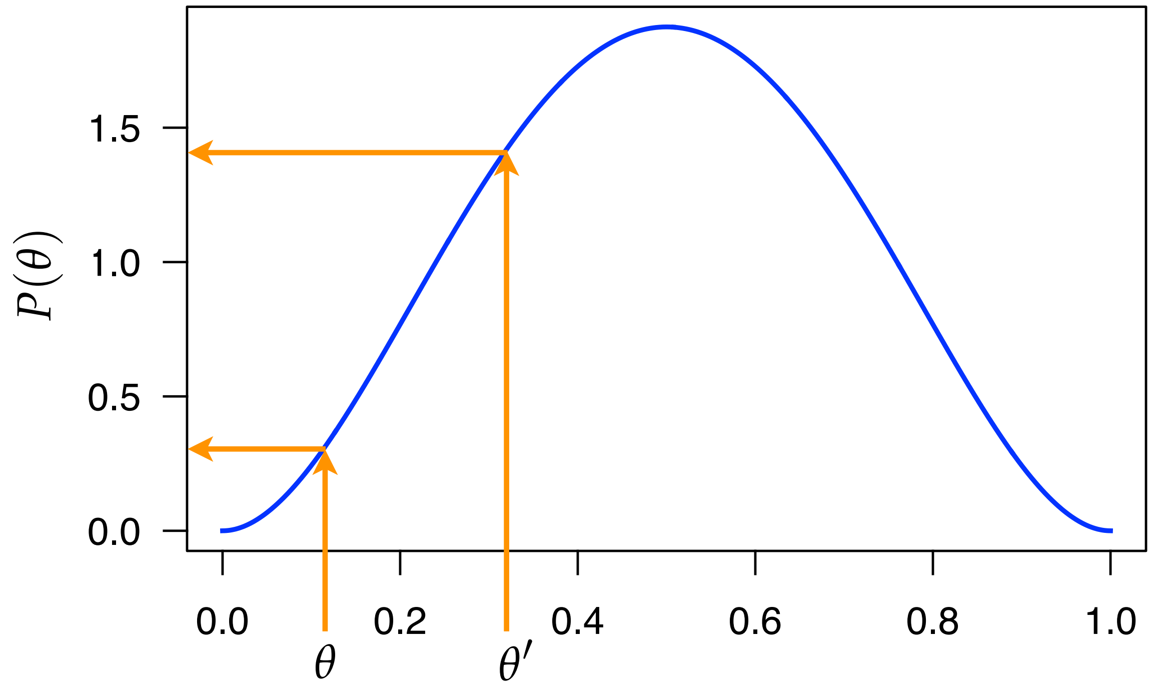
We can just look up the prior probability of a given parameter value



Approximating the Joint Posterior Probability Density using MCMC

The prior for each parameter is specified

We can just look up the prior probability of a given parameter value



Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?

$$R = \min\left[1, \boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)}} \times \boxed{\frac{\Pr(\theta')}{\Pr(\theta)}} \times \boxed{\frac{\Pr(\theta' \rightarrow \theta)}{\Pr(\theta \rightarrow \theta')}}\right]$$

likelihood ratio   prior ratio   proposal ratio

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - How do we calculate the proposal ratio?

$$R = \min\left[1, \underbrace{\boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)}}}_{\text{likelihood ratio}} \times \underbrace{\boxed{\frac{\Pr(\theta')}{\Pr(\theta)}}}_{\text{prior ratio}} \times \underbrace{\boxed{\frac{\Pr(\theta' \rightarrow \theta)}{\Pr(\theta \rightarrow \theta')}}}_{\text{proposal ratio}}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - How do we calculate the proposal ratio?

$$R = \min\left[1, \underbrace{\boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)}}}_{\text{likelihood ratio}} \times \underbrace{\boxed{\frac{\Pr(\theta')}{\Pr(\theta)}}}_{\text{prior ratio}}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - How do we calculate the proposal ratio?

$$P(\theta' \mid X) \propto \boxed{P(X \mid \theta')P(\theta')}$$

$$R = \min\left[1, \frac{\mathrm{Pr}(X \mid \theta')}{\mathrm{Pr}(X \mid \theta)} \times \frac{\mathrm{Pr}(\theta')}{\mathrm{Pr}(\theta)}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - How do we calculate the proposal ratio?

$$P(\theta' \mid X) \propto P(X \mid \theta')P(\theta')$$

$$R = \min\left[1, \frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right]$$

$$P(\theta \mid X) \propto \boxed{P(X \mid \theta)P(\theta)}$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - How do we calculate the proposal ratio?

$$R = \min\left[1, \boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}}\right]$$

$$\frac{P(\theta' \mid X)}{P(\theta \mid X)} = \frac{P(X \mid \theta')P(\theta')}{P(X \mid \theta)P(\theta)}$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
   - How do we calculate the likelihood for a given parameter value, θ?
   - How do we calculate the prior for a given parameter value, θ?
   - That means we can explore the posterior probability density without having to compute the marginal likelihood!!

$$R = \min\left[1, \boxed{\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}}\right]$$

$$\frac{P(\theta' \mid X)}{P(\theta \mid X)} = \frac{P(X \mid \theta')P(\theta')}{P(X \mid \theta)P(\theta)}$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change

$$R = \min\left[1, \frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \dots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
5. Generate a uniform random variable, $u \sim \text{Uniform}(0,1)$, accept if $u < R$

$$R = \min\left[1, \frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

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

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

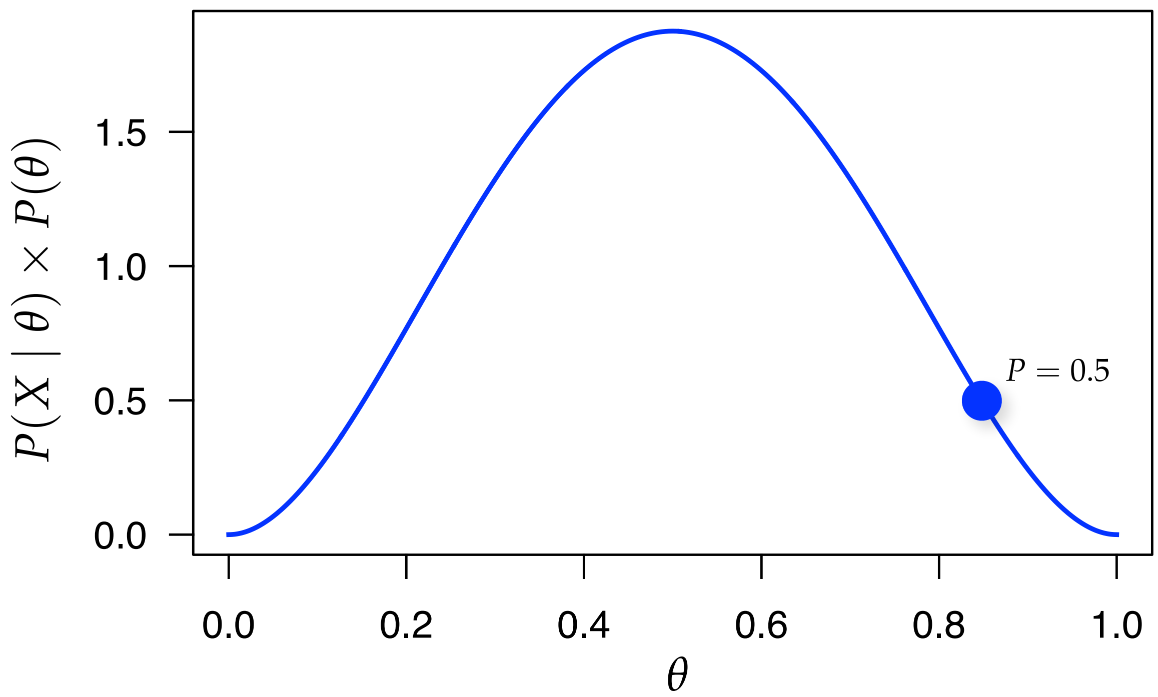
1. Initialize the chain with some random values for all parameters, including the tree with branch lengths, $\Theta = \{\tau, \nu, \pi, \ldots\}$
2. Select a parameter, θ, to update (alter) according to the proposal probabilities
3. Propose a new value, θ', for the selected parameter via the proposal mechanism
4. Calculate the probability of accepting the proposed change
5. Generate a uniform random variable, $u \sim$ Uniform$(0,1)$, accept if $u < R$
6. Repeat steps 2–5 an 'adequate' number of times

$$R = \min\left[1, \frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

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

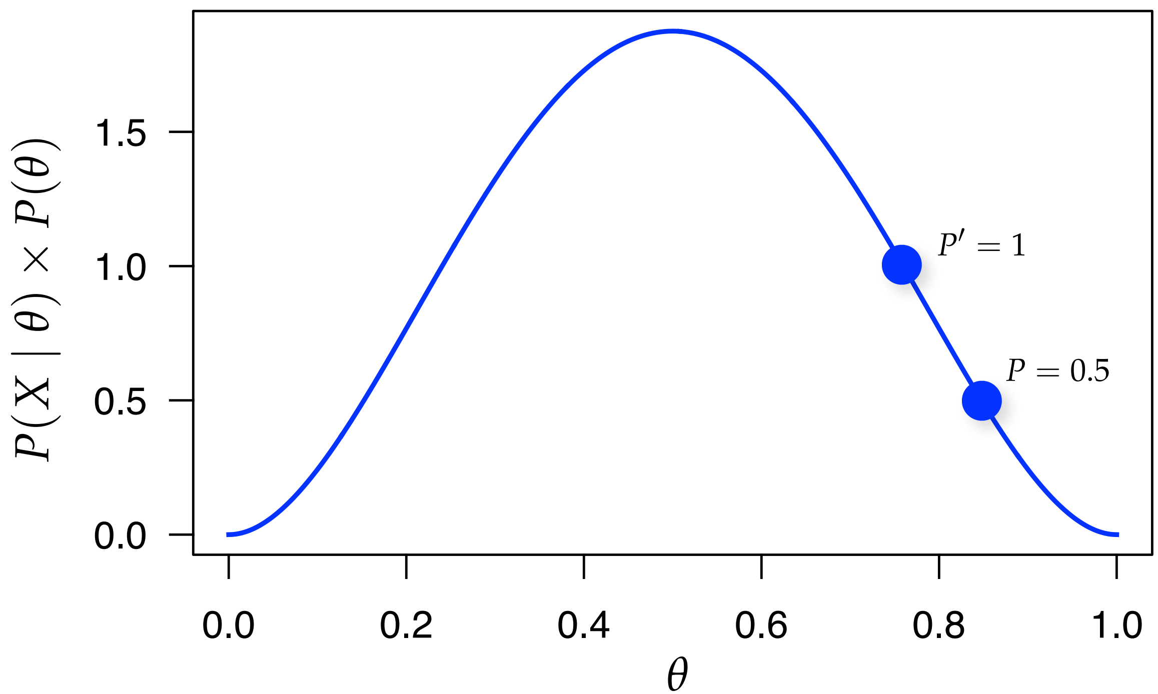
Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm



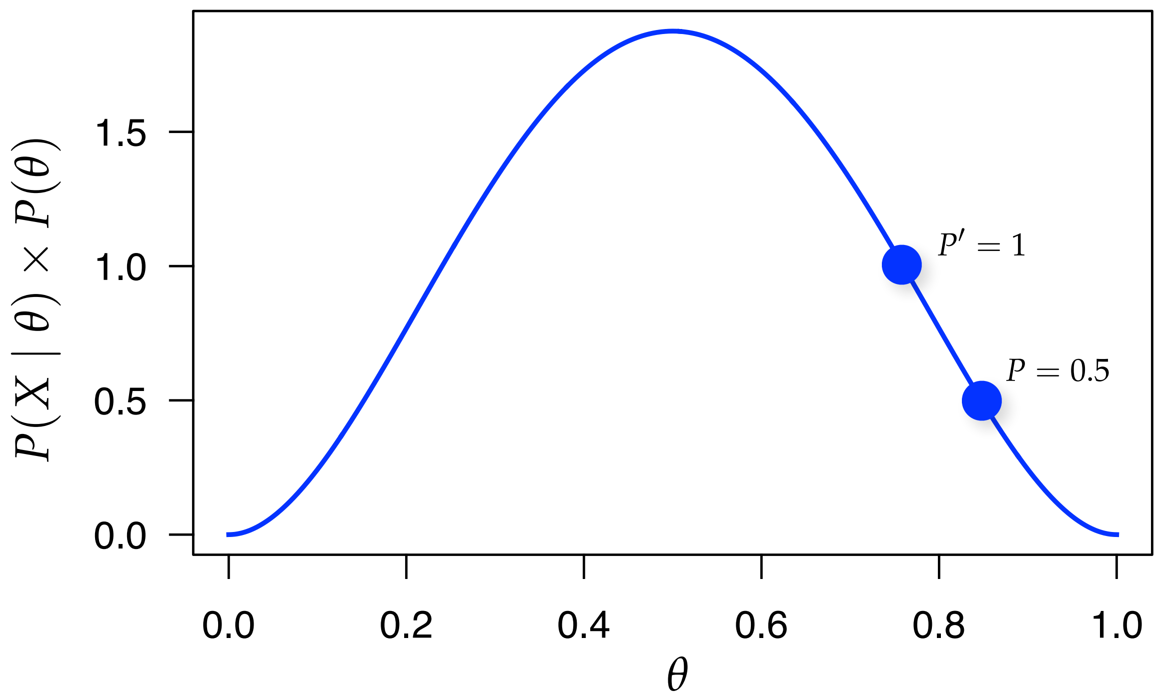
Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm



Approximating the Joint Posterior Probability Density using MCMC

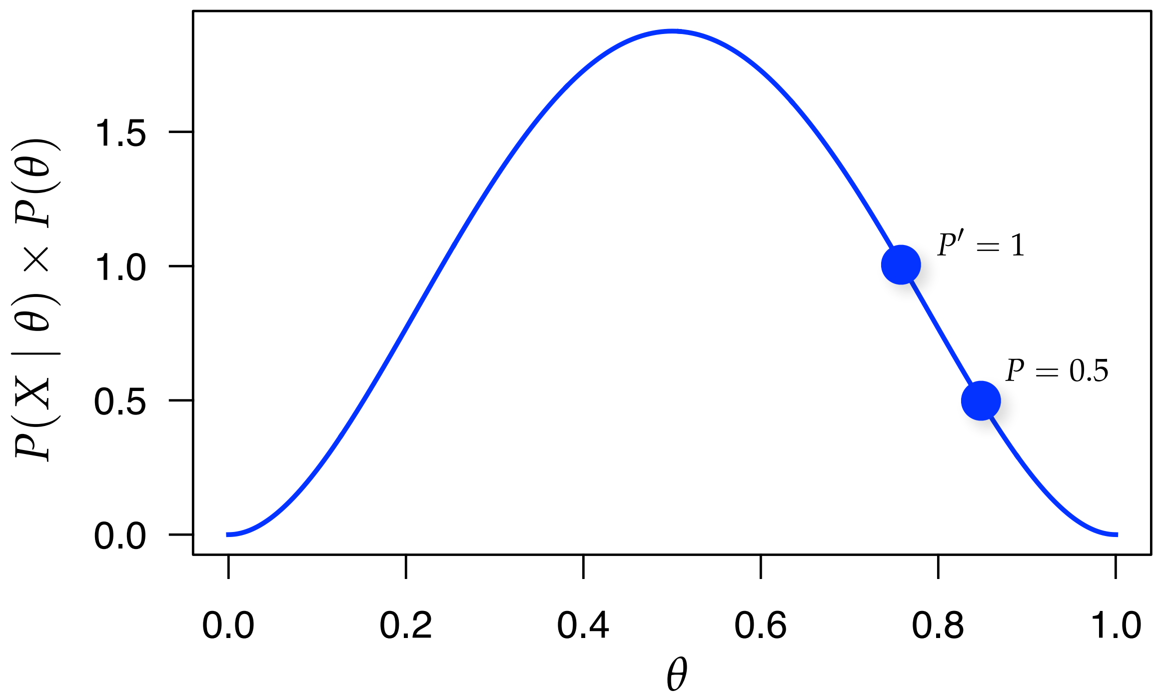
The Metropolis-Hastings algorithm



$$R = \min\left[1, \frac{1}{0.5}\right] = 1$$

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

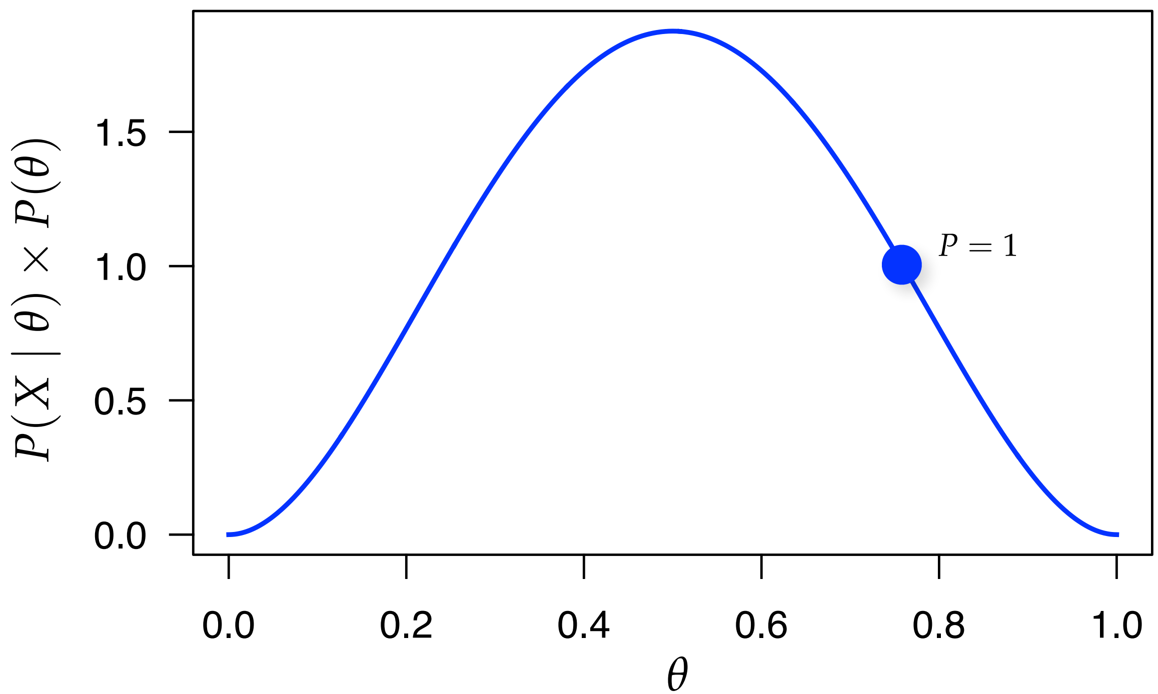


$$R = \min\left[1, \frac{1}{0.5}\right] = 1$$

$$u \sim \text{Uniform}(0, 1), \quad u = 0.983$$

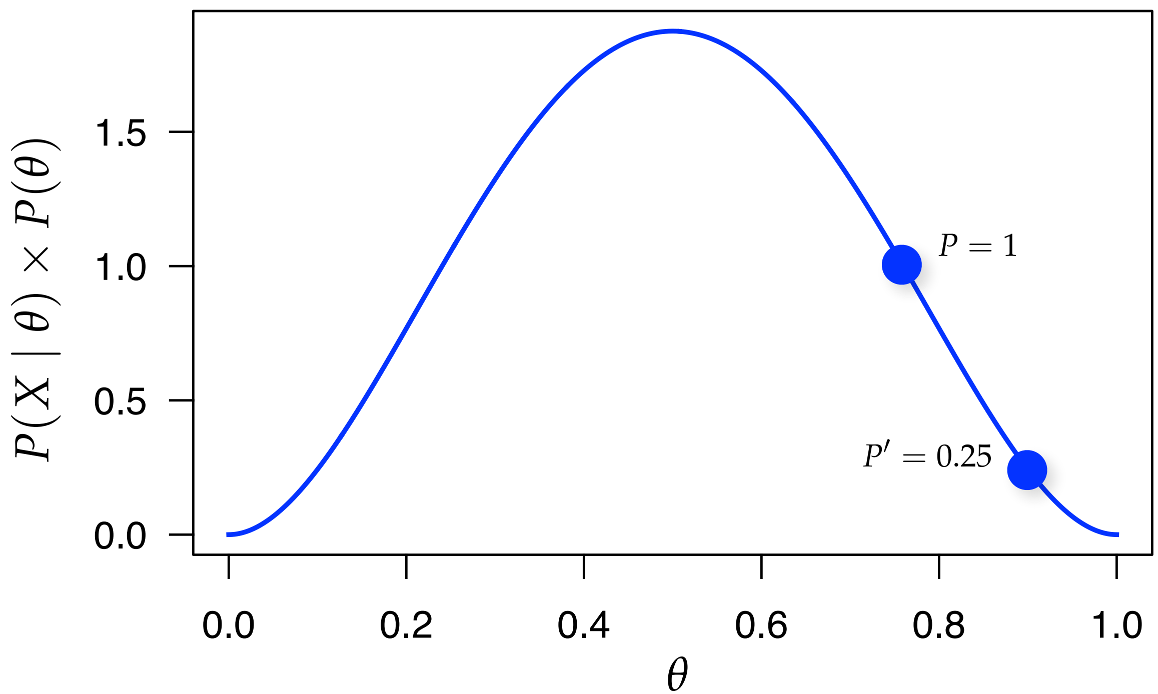
Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm



Approximating the Joint Posterior Probability Density using MCMC

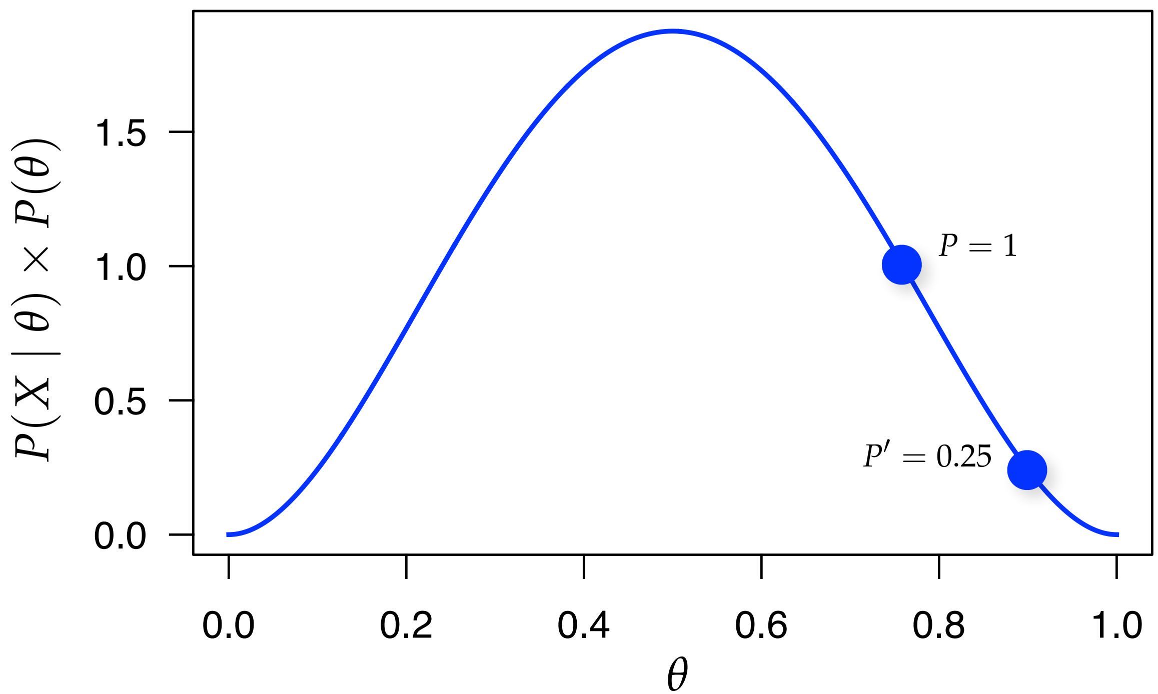
The Metropolis-Hastings algorithm



$$R = \min\left[1, \frac{0.25}{1}\right] = 0.25$$

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

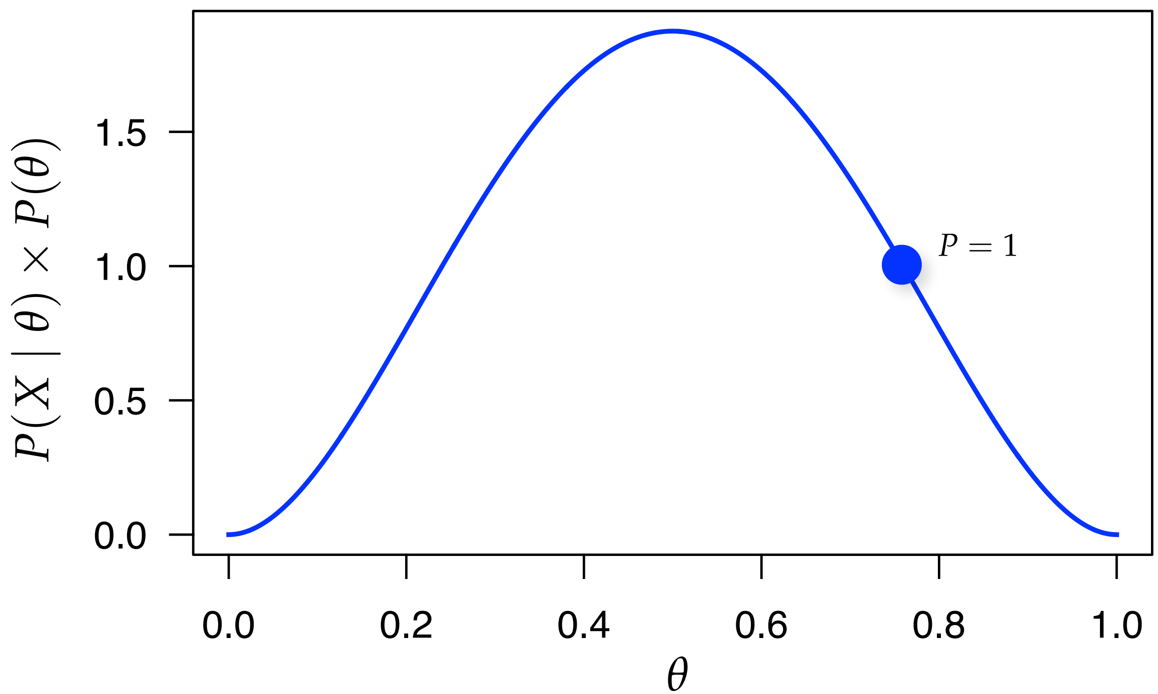


$$R = \min\left[1, \frac{0.25}{1}\right] = 0.25$$

$$u \sim \text{Uniform}(0,1), \quad u = 0.261$$

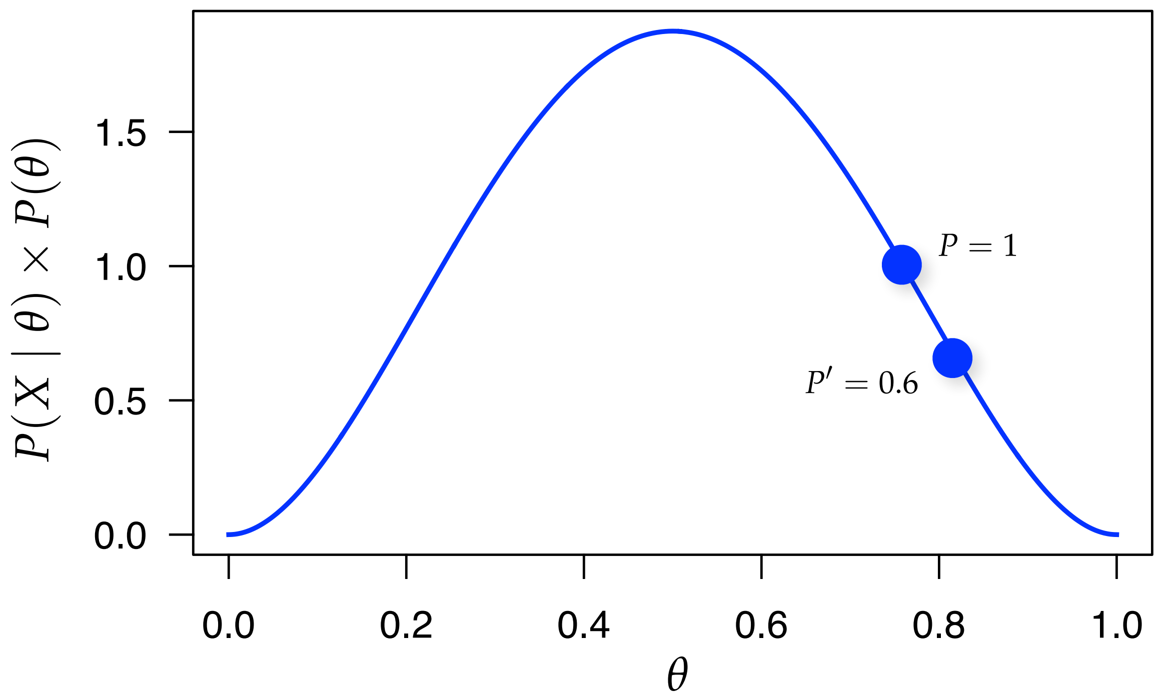
Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm



Approximating the Joint Posterior Probability Density using MCMC

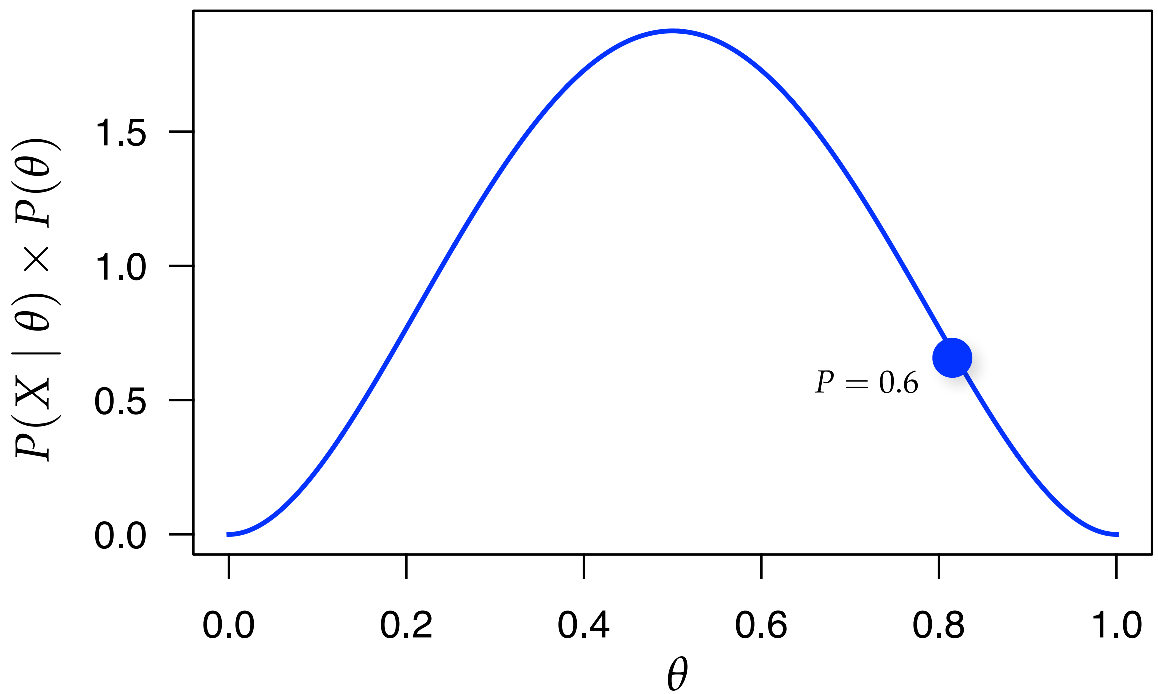
The Metropolis-Hastings algorithm



$$R = \min\left[1, \frac{0.6}{1}\right] = 0.6$$

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm

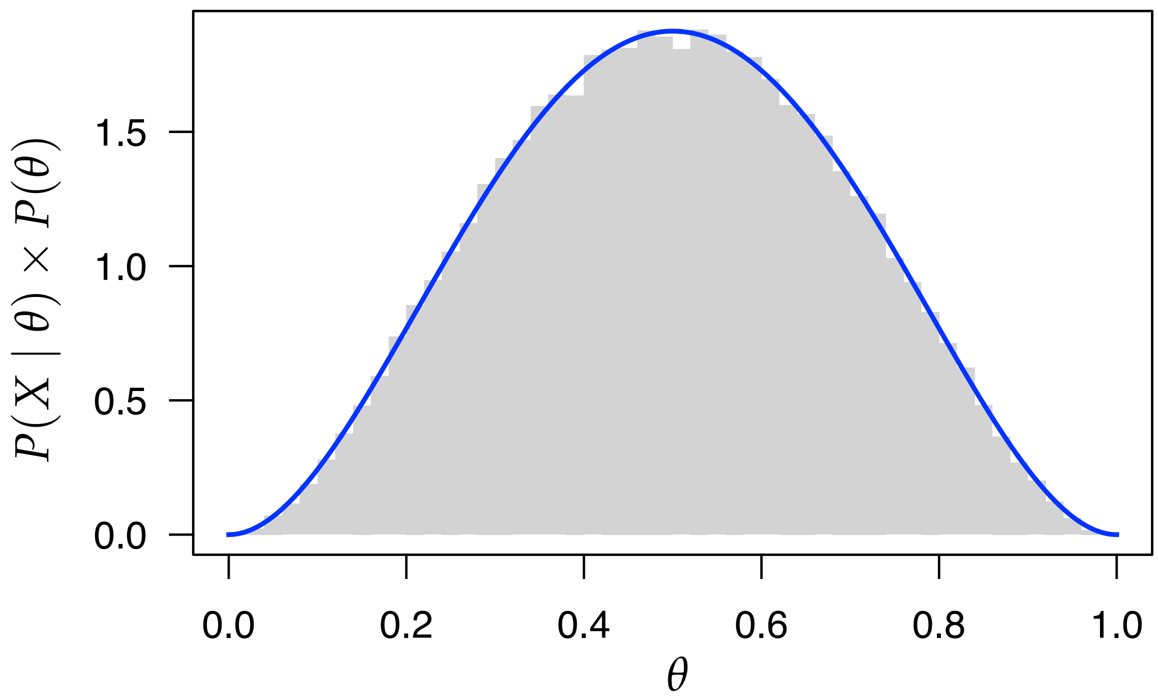


$$R = \min\left[1, \frac{0.6}{1}\right] = 0.6$$

$$u \sim \text{Uniform}(0, 1), \quad u = 0.128$$

Approximating the Joint Posterior Probability Density using MCMC

The Metropolis-Hastings algorithm



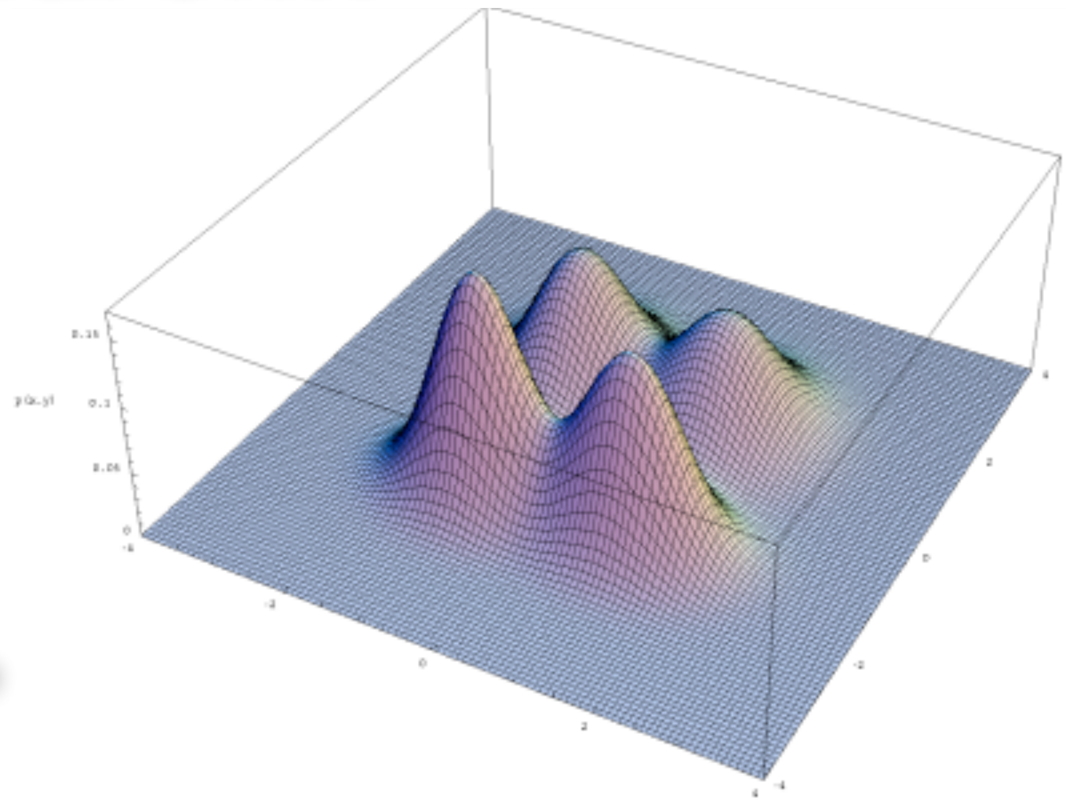
Approximating the Joint Posterior Probability Density using MCMC

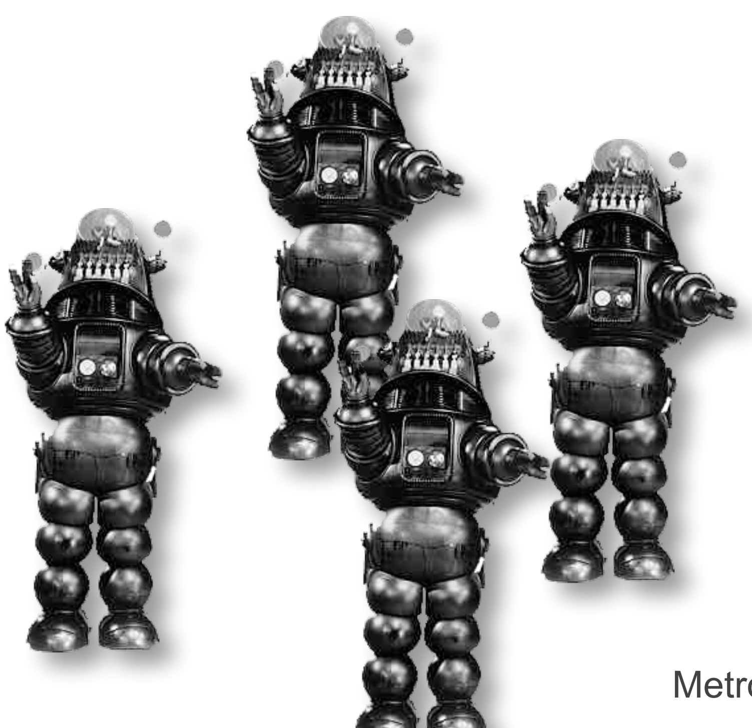
MCMCRobot demo!

https://plewis.github.io/applets/mcmc-robot/

Approximating the Joint Posterior Probability Density using MCMC

Robot Squadron!!





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

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature
chain	0.25
0	
1	
2	
3	

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature
chain	0.25
0	1.00
1	
2	
3	

$\beta_0 = 1/(1 + 0 \cdot 0.25)$

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2		
3		

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2	0.67	$\beta_2 = 1/(1 + 2 \cdot 0.25)$
3		

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
    - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

chain	temperature 0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2	0.67	$\beta_2 = 1/(1 + 2 \cdot 0.25)$
3	0.57	$\beta_3 = 1/(1 + 3 \cdot 0.25)$

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2	0.67	$\beta_2 = 1/(1 + 2 \cdot 0.25)$
3	0.57	$\beta_3 = 1/(1 + 3 \cdot 0.25)$

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2	0.67	$\beta_2 = 1/(1 + 2 \cdot 0.25)$
3	0.57	$\beta_3 = 1/(1 + 3 \cdot 0.25)$

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \rightarrow \theta)}{\Pr(\theta \rightarrow \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	
0	1.00	$\beta_0 = 1/(1 + 0 \cdot 0.25)$
1	0.80	$\beta_1 = 1/(1 + 1 \cdot 0.25)$
2	0.67	$\beta_2 = 1/(1 + 2 \cdot 0.25)$
3	0.57	$\beta_3 = 1/(1 + 3 \cdot 0.25)$

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.
   - the degree of incremental heating is controlled by the temperature parameter, T.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature	
chain	0.25	0.20
0	1.00	1.00
1	0.80	0.83
2	0.67	0.71
3	0.57	0.63

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \rightarrow \theta)}{\Pr(\theta \rightarrow \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.
   - the degree of incremental heating is controlled by the temperature parameter, T.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature		
chain	0.25	0.20	0.15
0	1.00	1.00	1.00
1	0.80	0.83	0.87
2	0.67	0.71	0.77
3	0.57	0.63	0.69

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.
   - the degree of incremental heating is controlled by the temperature parameter, T.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature			
chain	0.25	0.20	0.15	0.10
0	1.00	1.00	1.00	1.00
1	0.80	0.83	0.87	0.91
2	0.67	0.71	0.77	0.83
3	0.57	0.63	0.69	0.77

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \to \theta)}{\Pr(\theta \to \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.
   - the degree of incremental heating is controlled by the temperature parameter, T.

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$

	temperature				
	chain	0.25	0.20	0.15	0.10
cold chain	0	1.00	1.00	1.00	1.00
	1	0.80	0.83	0.87	0.91
	2	0.67	0.71	0.77	0.83
	3	0.57	0.63	0.69	0.77

   - the incremental heating successively 'flattens' the posterior visited by each chain by making the acceptance probability of the i^{th} chain more 'permissive':

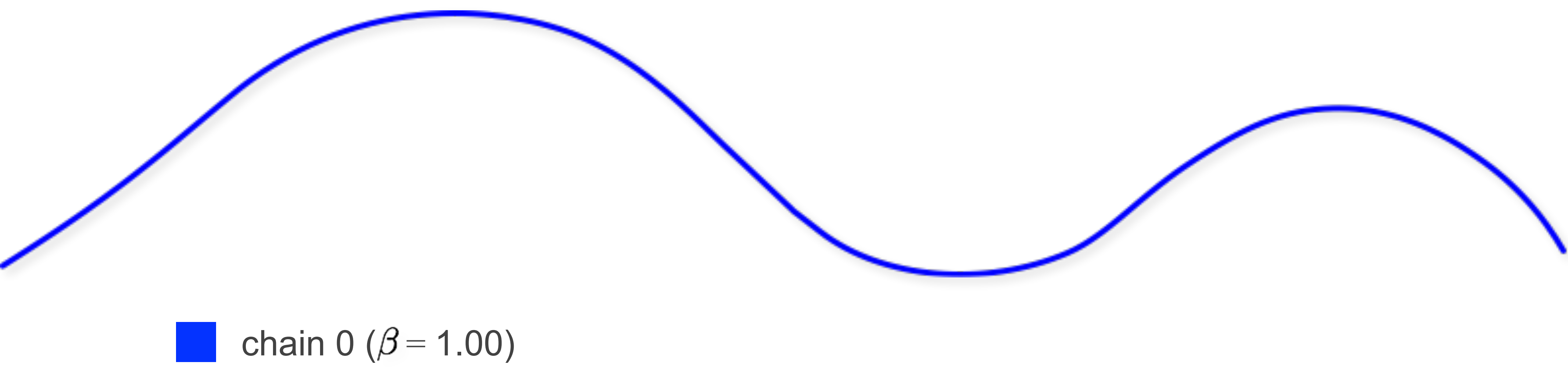
$$R_i = \min\left[1, \left(\frac{\Pr(X \mid \theta')}{\Pr(X \mid \theta)} \times \frac{\Pr(\theta')}{\Pr(\theta)}\right)^{\beta_i} \times \frac{\Pr(\theta' \rightarrow \theta)}{\Pr(\theta \rightarrow \theta')}\right]$$

   - this allows heated chains to more readily traverse regions of low probability.
   - the degree of incremental heating is controlled by the temperature parameter, T.
   - samples are only collected by the 'cold' chain (*i.e.*, the undistorted posterior).

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

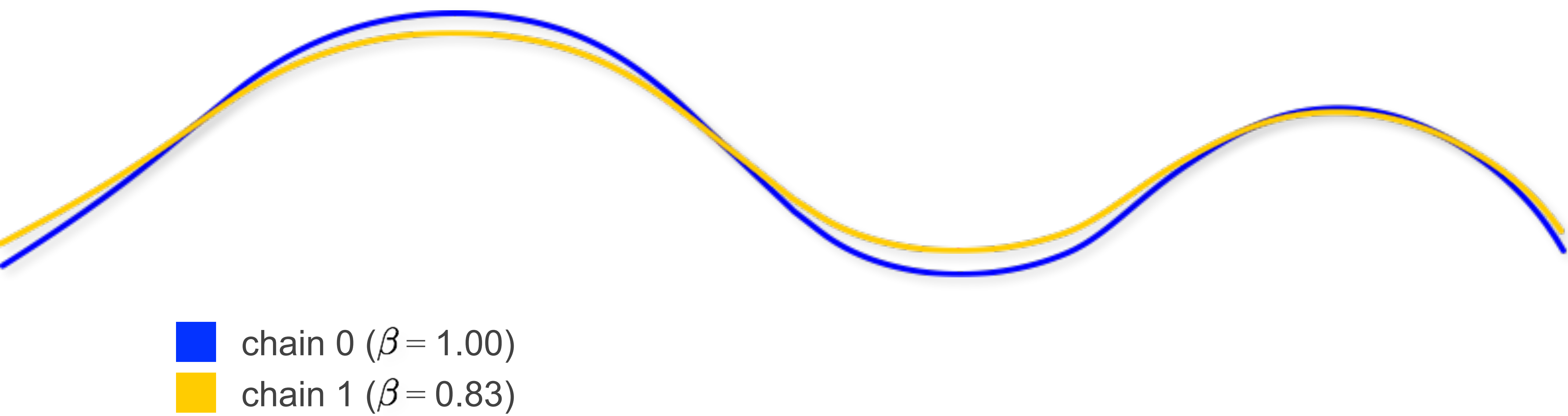
1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
   - the cold chain samples the true posterior, whereas the heated chains sample successively ‘flattened’ distortions of the posterior



Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC[3] algorithm

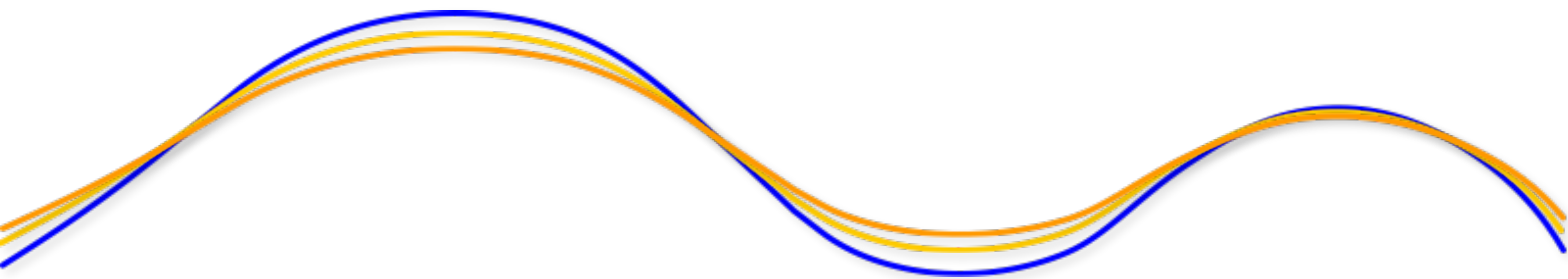
1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
   - the cold chain samples the true posterior, whereas the heated chains sample successively 'flattened' distortions of the posterior



Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
   - the cold chain samples the true posterior, whereas the heated chains sample successively 'flattened' distortions of the posterior

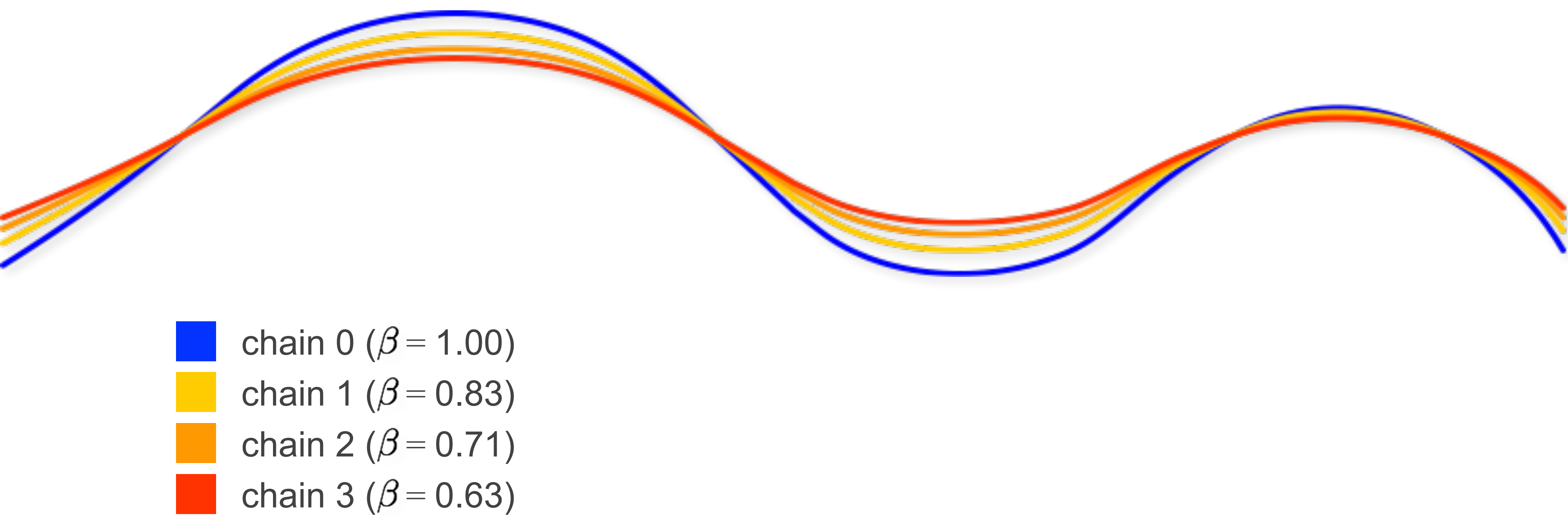


chain 0 ($\beta = 1.00$)
chain 1 ($\beta = 0.83$)
chain 2 ($\beta = 0.71$)

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

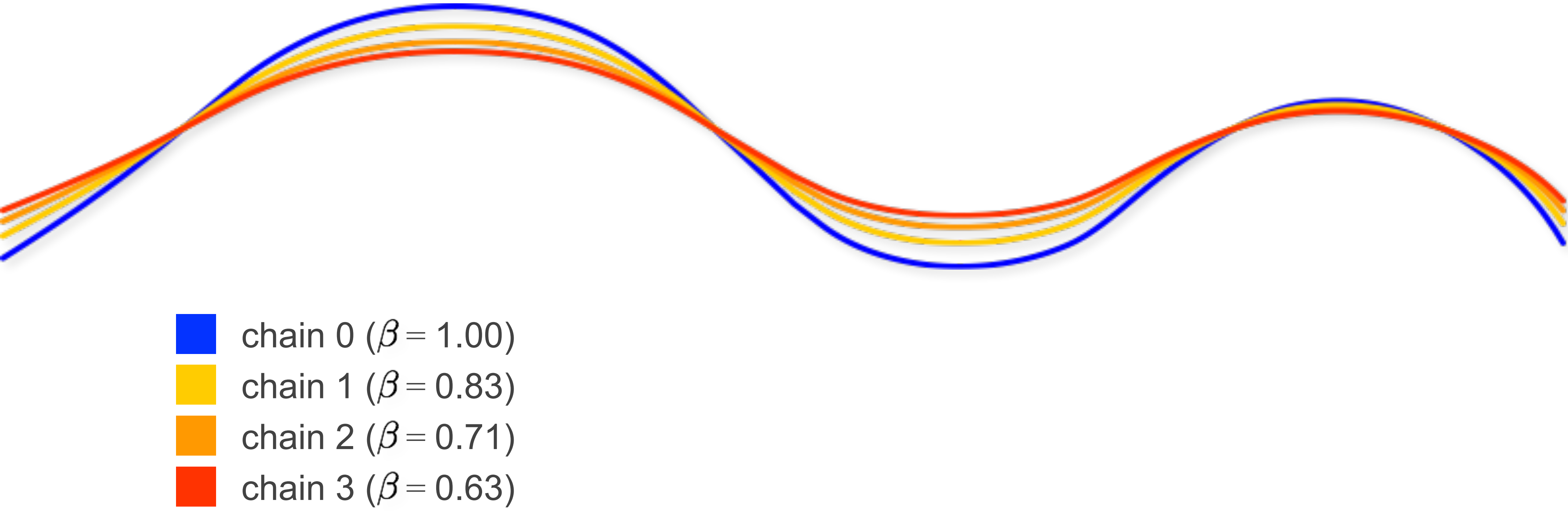
1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
    - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
    - the cold chain samples the true posterior, whereas the heated chains sample successively 'flattened' distortions of the posterior



Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

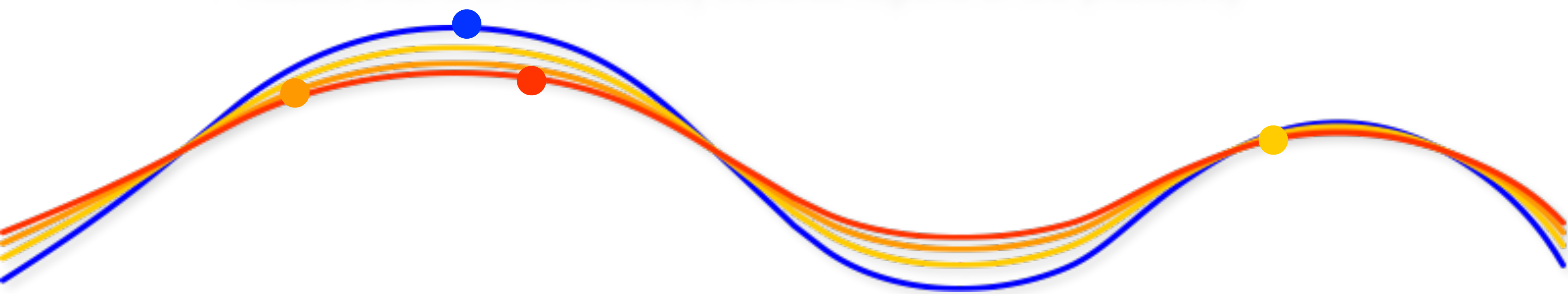
1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
    - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
    - the cold chain samples the true posterior, whereas the heated chains sample successively ‘flattened’ distortions of the posterior
    - heated chains to more readily traverse regions of low probability



Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
   - posterior of chain i is raised to a power, β_i: the heat of chain $i = 1/(1 + iT)$
   - the cold chain samples the true posterior, whereas the heated chains sample successively 'flattened' distortions of the posterior
   - heated chains to more readily traverse regions of low probability

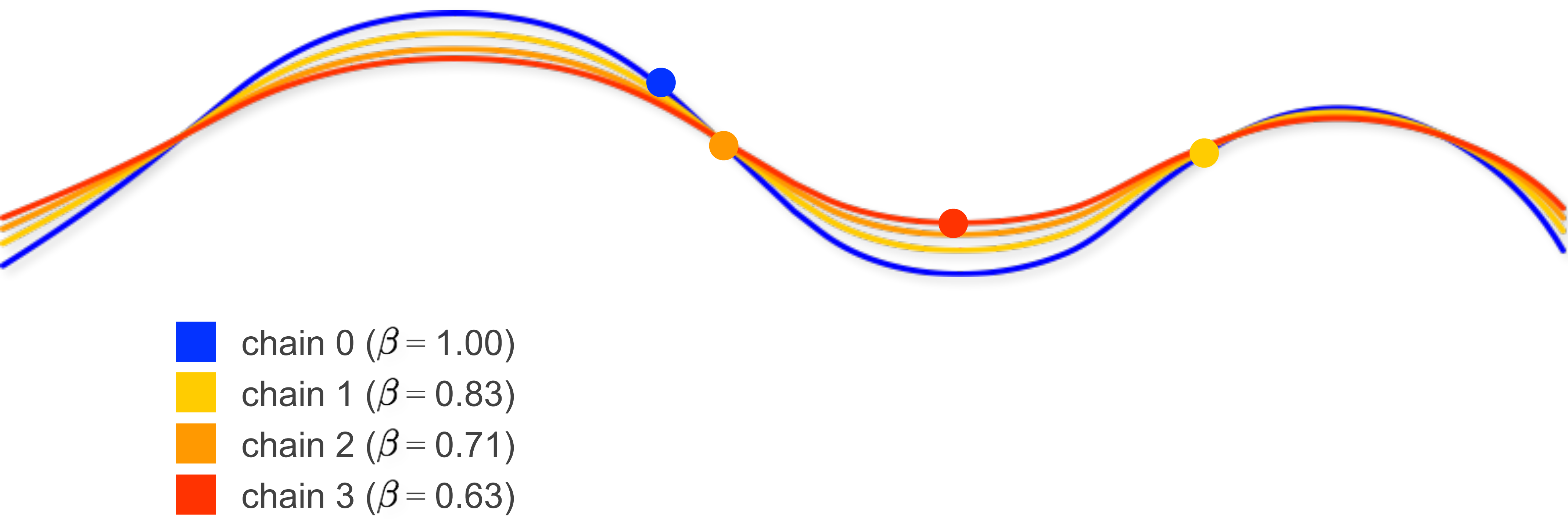


- chain 0 ($\beta = 1.00$)
- chain 1 ($\beta = 0.83$)
- chain 2 ($\beta = 0.71$)
- chain 3 ($\beta = 0.63$)

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
3. At prescribed intervals, two chains are randomly selected to swap.

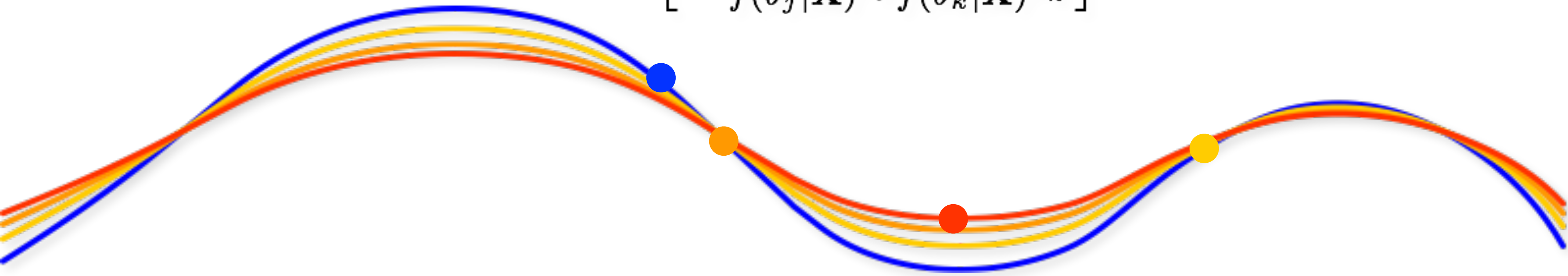


Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
3. At prescribed intervals, two chains are randomly selected to swap.
   - we compute the acceptance probability of swapping the two chains.

$$R = \min\left[1, \frac{f(\theta_k|\mathbf{X})^{\beta_j} f(\theta_j|\mathbf{X})^{\beta_k}}{f(\theta_j|\mathbf{X})^{\beta_j} f(\theta_k|\mathbf{X})^{\beta_k}}\right]$$

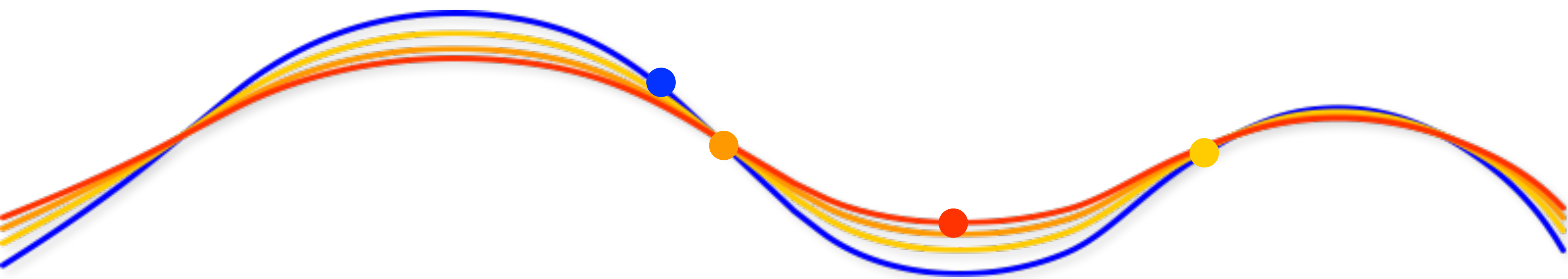


- chain 0 ($\beta = 1.00$)
- chain 1 ($\beta = 0.83$)
- chain 2 ($\beta = 0.71$)
- chain 3 ($\beta = 0.63$)

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC³ algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
3. At prescribed intervals, two chains are randomly selected to swap.
   - we compute the acceptance probability of swapping the two chains.
   - if accepted, the chains swap positions (and in computer memory)

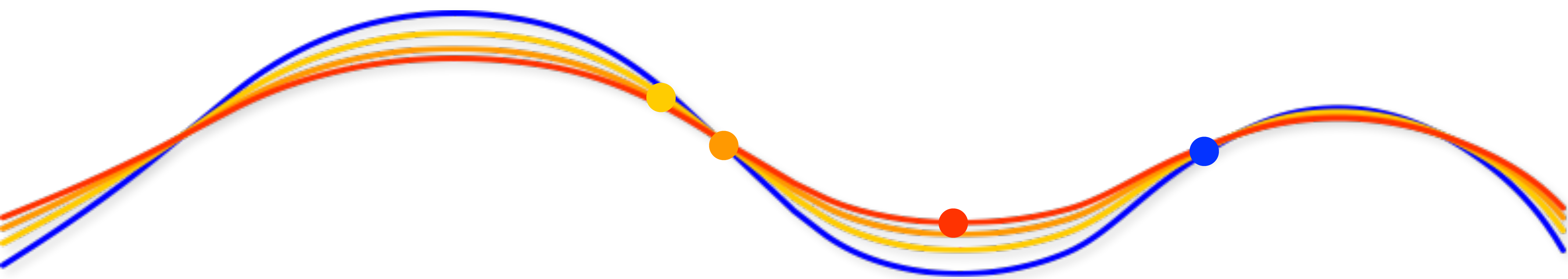


- chain 0 ($\beta = 1.00$)
- chain 1 ($\beta = 0.83$)
- chain 2 ($\beta = 0.71$)
- chain 3 ($\beta = 0.63$)

Approximating the Joint Posterior Probability Density using Metropolis-Coupled MCMC

The MC^3 algorithm

1. Initialize N independent M-H MCMC chains with random values for all parameters.
2. The chains are incrementally heated, such that the first chain is cold (unmodified).
3. At prescribed intervals, two chains are randomly selected to swap.
4. Only samples from the cold chain are used to approximate the posterior.



chain 0 ($\beta = 1.00$)
chain 1 ($\beta = 0.83$)
chain 2 ($\beta = 0.71$)
chain 3 ($\beta = 0.63$)

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

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

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

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

- *e.g.*, the frequency of a sampled clade provides an estimate of its nodal probability

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

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

- *e.g.*, the frequency of a sampled clade provides an estimate of its nodal probability

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

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

Each row in our log file—with values of all model parameters—is a sample from the *joint* posterior probability density.

[ID: 2325481386]

Gen	LnL	TL	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)	alpha
1	-13413.769	1.313	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.250000	0.250000	0.250000	0.250000	0.500000
1000	-10429.772	0.904	0.100364	0.271178	0.057126	0.095681	0.404818	0.070833	0.276201	0.173231	0.228359	0.322209	0.845634
2000	-10420.654	0.980	0.115937	0.254216	0.041309	0.051039	0.455344	0.082157	0.291050	0.181003	0.231042	0.296904	0.670406
3000	-10417.930	0.961	0.137253	0.264348	0.037891	0.056962	0.426295	0.077251	0.291050	0.181003	0.231042	0.296904	0.901480
4000	-10423.816	0.925	0.101065	0.273786	0.035266	0.067623	0.441301	0.080958	0.290603	0.185952	0.231800	0.291644	0.859284
5000	-10425.264	1.002	0.135985	0.259584	0.048509	0.057733	0.430436	0.067753	0.289106	0.189615	0.210373	0.310906	0.671675
6000	-10421.366	0.962	0.119016	0.268203	0.041284	0.062913	0.415543	0.093041	0.281133	0.187367	0.234148	0.297353	0.824395
7000	-10417.840	0.981	0.123308	0.246185	0.032588	0.070686	0.443381	0.083851	0.298478	0.186125	0.221560	0.293837	0.644508
8000	-10420.174	1.058	0.129152	0.263612	0.036846	0.061359	0.424323	0.084708	0.284539	0.192084	0.216456	0.306921	0.691606
9000	-10419.701	0.980	0.101173	0.266573	0.035445	0.072158	0.438826	0.085825	0.285541	0.188378	0.229610	0.296471	0.687021
10000	-10423.917	1.015	0.100312	0.289851	0.045985	0.059364	0.422372	0.082115	0.285505	0.176257	0.228230	0.310007	0.684473
11000	-10418.487	0.945	0.107911	0.270677	0.049322	0.063833	0.421602	0.086655	0.279829	0.188085	0.233921	0.298165	0.860128
12000	-10420.169	0.893	0.115085	0.270950	0.038203	0.070506	0.417478	0.087778	0.288131	0.191473	0.231758	0.288638	0.723312
13000	-10419.081	0.922	0.115323	0.269076	0.036184	0.069919	0.429555	0.079943	0.294340	0.187665	0.227043	0.290952	0.784700
14000	-10423.817	1.030	0.112545	0.254842	0.042601	0.077867	0.436797	0.075348	0.283706	0.189549	0.224014	0.302731	0.615981
15000	-10424.879	0.944	0.131641	0.260134	0.043160	0.069779	0.421550	0.073736	0.296187	0.175620	0.219147	0.309046	0.797970
16000	-10426.143	0.940	0.117469	0.266011	0.056463	0.049593	0.441326	0.069139	0.282578	0.203117	0.231372	0.282933	0.792757
17000	-10421.133	0.978	0.134024	0.277374	0.040419	0.056384	0.416233	0.075565	0.289061	0.187968	0.225825	0.297145	0.767063
18000	-10418.290	0.930	0.104450	0.251683	0.041434	0.063649	0.455528	0.083256	0.287086	0.189510	0.226700	0.296704	0.767072
19000	-10420.052	0.972	0.121227	0.274901	0.037023	0.083743	0.414224	0.068881	0.289061	0.187968	0.225825	0.297145	0.758345
20000	-10425.127	0.955	0.099741	0.277386	0.043745	0.069447	0.433059	0.076622	0.292229	0.197483	0.212827	0.297461	0.645034
21000	-10421.087	0.939	0.105737	0.258514	0.039941	0.094773	0.429045	0.071991	0.292778	0.192129	0.217655	0.297438	0.692877
22000	-10421.805	0.926	0.111237	0.293260	0.047595	0.061320	0.409044	0.077544	0.286897	0.197795	0.222410	0.292899	0.797696
23000	-10422.326	0.943	0.123590	0.240213	0.047236	0.048864	0.453312	0.086786	0.291024	0.187438	0.225934	0.295603	0.851381
24000	-10417.974	0.938	0.123674	0.274369	0.051414	0.065387	0.413009	0.072146	0.291024	0.187438	0.225934	0.295603	0.801620
25000	-10422.454	0.996	0.132415	0.249036	0.036744	0.063052	0.457012	0.061741	0.299053	0.171847	0.226435	0.302665	0.607659
26000	-10424.506	0.892	0.122118	0.235061	0.042240	0.063788	0.462004	0.074790	0.302331	0.170502	0.220011	0.307156	0.812245
27000	-10420.001	0.953	0.128264	0.263415	0.040470	0.058989	0.432138	0.076724	0.279181	0.190422	0.234369	0.296028	0.824956

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

Each row in our log file—with values of all model parameters—is a sample from the *joint* posterior probability density.

[ID: 2325481386]

Gen	LnL	TL	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)	alpha
1	-13413.769	1.313	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.250000	0.250000	0.250000	0.250000	0.500000
1000	-10429.772	0.904	0.100364	0.271178	0.057126	0.095681	0.404818	0.070833	0.276201	0.173231	0.228359	0.322209	0.845634
2000	-10420.654	0.980	0.115937	0.254216	0.041309	0.051039	0.455344	0.082157	0.291050	0.181003	0.231042	0.296904	0.670406
3000	-10417.930	0.961	0.137253	0.264348	0.037891	0.056962	0.426295	0.077251	0.291050	0.181003	0.231042	0.296904	0.901480
4000	-10423.816	0.925	0.101065	0.273786	0.035266	0.067623	0.441301	0.080958	0.290603	0.185952	0.231800	0.291644	0.859284
5000	-10425.264	1.002	0.135985	0.259584	0.048509	0.057733	0.430436	0.067753	0.289106	0.189615	0.210373	0.310906	0.671675
6000	-10421.366	0.962	0.119016	0.268203	0.041284	0.062913	0.415543	0.093041	0.281133	0.187367	0.234148	0.297353	0.824395
7000	-10417.840	0.981	0.123308	0.246185	0.032588	0.070686	0.443381	0.083851	0.298478	0.186125	0.221560	0.293837	0.644508
8000	-10420.174	1.058	0.129152	0.263612	0.036846	0.061359	0.424323	0.084708	0.284539	0.192084	0.216456	0.306921	0.691606
9000	-10419.701	0.980	0.101173	0.266573	0.035445	0.072158	0.438826	0.085825	0.285541	0.188378	0.229610	0.296471	0.687021
10000	-10423.917	1.015	0.100312	0.289851	0.045985	0.059364	0.422372	0.082115	0.285505	0.176257	0.228230	0.310007	0.684473
11000	-10418.487	0.945	0.107911	0.270677	0.049322	0.063833	0.421602	0.086655	0.279829	0.188085	0.233921	0.298165	0.860128
12000	-10420.169	0.893	0.115085	0.270950	0.038203	0.070506	0.417478	0.087778	0.288131	0.191473	0.231758	0.288638	0.723312
13000	-10419.081	0.922	0.115323	0.269076	0.036184	0.069919	0.429555	0.079943	0.294340	0.187665	0.227043	0.290952	0.784700
14000	-10423.817	1.030	0.112545	0.254842	0.042601	0.077867	0.436797	0.075348	0.283706	0.189549	0.224014	0.302731	0.615981
15000	-10424.879	0.944	0.131641	0.260134	0.043160	0.069779	0.421550	0.073736	0.296187	0.175620	0.219147	0.309046	0.797970
16000	-10426.143	0.940	0.117469	0.266011	0.056463	0.049593	0.441326	0.069139	0.282578	0.203117	0.231372	0.282933	0.792757
17000	-10421.133	0.978	0.134024	0.277374	0.040419	0.056384	0.416233	0.075565	0.289061	0.187968	0.225825	0.297145	0.767063
18000	-10418.290	0.930	0.104450	0.251683	0.041434	0.063649	0.455528	0.083256	0.287086	0.189510	0.226700	0.296704	0.767072
19000	-10420.052	0.972	0.121227	0.274901	0.037023	0.083743	0.414224	0.068881	0.289061	0.187968	0.225825	0.297145	0.758345
20000	-10425.127	0.955	0.099741	0.277386	0.043745	0.069447	0.433059	0.076622	0.292229	0.197483	0.212827	0.297461	0.645034
21000	-10421.087	0.939	0.105737	0.258514	0.039941	0.094773	0.429045	0.071991	0.292778	0.192129	0.217655	0.297438	0.692877
22000	-10421.805	0.926	0.111237	0.293260	0.047595	0.061320	0.409044	0.077544	0.286897	0.197795	0.222410	0.292899	0.797696
23000	-10422.326	0.943	0.123590	0.240213	0.047236	0.048864	0.453312	0.086786	0.291024	0.187438	0.225934	0.295603	0.851381
24000	-10417.974	0.938	0.123674	0.274369	0.051414	0.065387	0.413009	0.072146	0.291024	0.187438	0.225934	0.295603	0.801620
25000	-10422.454	0.996	0.132415	0.249036	0.036744	0.063052	0.457012	0.061741	0.299053	0.171847	0.226435	0.302665	0.607659
26000	-10424.506	0.892	0.122118	0.235061	0.042240	0.063788	0.462004	0.074790	0.302331	0.170502	0.220011	0.307156	0.812245
27000	-10420.001	0.953	0.128264	0.263415	0.040470	0.058989	0.432138	0.076724	0.279181	0.190422	0.234369	0.296028	0.824956

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

Each row in our log file—with values of all model parameters—is a sample from the *joint* posterior probability density.

[ID: 2325481386]

Gen	LnL	TL	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)	alpha
1	-13413.769	1.313	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.250000	0.250000	0.250000	0.250000	0.500000
1000	-10429.772	0.904	0.100364	0.271178	0.057126	0.095681	0.404818	0.070833	0.276201	0.173231	0.228359	0.322209	0.845634
2000	-10420.654	0.980	0.115937	0.254216	0.041309	0.051039	0.455344	0.082157	0.291050	0.181003	0.231042	0.296904	0.670406
3000	-10417.930	0.961	0.137253	0.264348	0.037891	0.056962	0.426295	0.077251	0.291050	0.181003	0.231042	0.296904	0.901480
4000	-10423.816	0.925	0.101065	0.273786	0.035266	0.067623	0.441301	0.080958	0.290603	0.185952	0.231800	0.291644	0.859284
5000	-10425.264	1.002	0.135985	0.259584	0.048509	0.057733	0.430436	0.067753	0.289106	0.189615	0.210373	0.310906	0.671675
6000	-10421.366	0.962	0.119016	0.268203	0.041284	0.062913	0.415543	0.093041	0.281133	0.187367	0.234148	0.297353	0.824395
7000	-10417.840	0.981	0.123308	0.246185	0.032588	0.070686	0.443381	0.083851	0.298478	0.186125	0.221560	0.293837	0.644508
8000	-10420.174	1.058	0.129152	0.263612	0.036846	0.061359	0.424323	0.084708	0.284539	0.192084	0.216456	0.306921	0.691606
9000	-10419.701	0.980	0.101173	0.266573	0.035445	0.072158	0.438826	0.085825	0.285541	0.188378	0.229610	0.296471	0.687021
10000	-10423.917	1.015	0.100312	0.289851	0.045985	0.059364	0.422372	0.082115	0.285505	0.176257	0.228230	0.310007	0.684473
11000	-10418.487	0.945	0.107911	0.270677	0.049322	0.063833	0.421602	0.086655	0.279829	0.188085	0.233921	0.298165	0.860128
12000	-10420.169	0.893	0.115085	0.270950	0.038203	0.070506	0.417478	0.087778	0.288131	0.191473	0.231758	0.288638	0.723312
13000	-10419.081	0.922	0.115323	0.269076	0.036184	0.069919	0.429555	0.079943	0.294340	0.187665	0.227043	0.290952	0.784700
14000	-10423.817	1.030	0.112545	0.254842	0.042601	0.077867	0.436797	0.075348	0.283706	0.189549	0.224014	0.302731	0.615981
15000	-10424.879	0.944	0.131641	0.260134	0.043160	0.069779	0.421550	0.073736	0.296187	0.175620	0.219147	0.309046	0.797970
16000	-10426.143	0.940	0.117469	0.266011	0.056463	0.049593	0.441326	0.069139	0.282578	0.203117	0.231372	0.282933	0.792757
17000	-10421.133	0.978	0.134024	0.277374	0.040419	0.056384	0.416233	0.075565	0.289061	0.187968	0.225825	0.297145	0.767063
18000	-10418.290	0.930	0.104450	0.251683	0.041434	0.063649	0.455528	0.083256	0.287086	0.189510	0.226700	0.296704	0.767072
19000	-10420.052	0.972	0.121227	0.274901	0.037023	0.083743	0.414224	0.068881	0.289061	0.187968	0.225825	0.297145	0.758345
20000	-10425.127	0.955	0.099741	0.277386	0.043745	0.069447	0.433059	0.076622	0.292229	0.197483	0.212827	0.297461	0.645034
21000	-10421.087	0.939	0.105737	0.258514	0.039941	0.094773	0.429045	0.071991	0.292778	0.192129	0.217655	0.297438	0.692877
22000	-10421.805	0.926	0.111237	0.293260	0.047595	0.061320	0.409044	0.077544	0.286897	0.197795	0.222410	0.292899	0.797696
23000	-10422.326	0.943	0.123590	0.240213	0.047236	0.048864	0.453312	0.086786	0.291024	0.187438	0.225934	0.295603	0.851381
24000	-10417.974	0.938	0.123674	0.274369	0.051414	0.065387	0.413009	0.072146	0.291024	0.187438	0.225934	0.295603	0.801620
25000	-10422.454	0.996	0.132415	0.249036	0.036744	0.063052	0.457012	0.061741	0.299053	0.171847	0.226435	0.302665	0.607659
26000	-10424.506	0.892	0.122118	0.235061	0.042240	0.063788	0.462004	0.074790	0.302331	0.170502	0.220011	0.307156	0.812245
27000	-10420.001	0.953	0.128264	0.263415	0.040470	0.058989	0.432138	0.076724	0.279181	0.190422	0.234369	0.296028	0.824956

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

Each column in our log file—with values for a single model parameter—is a sample from the *marginal* posterior probability density.

[ID: 2325481386]

Gen	LnL	TL	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)	alpha
1	-13413.769	1.313	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.250000	0.250000	0.250000	0.250000	0.500000
1000	-10429.772	0.904	0.100364	0.271178	0.057126	0.095681	0.404818	0.070833	0.276201	0.173231	0.228359	0.322209	0.845634
2000	-10420.654	0.980	0.115937	0.254216	0.041309	0.051039	0.455344	0.082157	0.291050	0.181003	0.231042	0.296904	0.670406
3000	-10417.930	0.961	0.137253	0.264348	0.037891	0.056962	0.426295	0.077251	0.291050	0.181003	0.231042	0.296904	0.901480
4000	-10423.816	0.925	0.101065	0.273786	0.035266	0.067623	0.441301	0.080958	0.290603	0.185952	0.231800	0.291644	0.859284
5000	-10425.264	1.002	0.135985	0.259584	0.048509	0.057733	0.430436	0.067753	0.289106	0.189615	0.210373	0.310906	0.671675
6000	-10421.366	0.962	0.119016	0.268203	0.041284	0.062913	0.415543	0.093041	0.281133	0.187367	0.234148	0.297353	0.824395
7000	-10417.840	0.981	0.123308	0.246185	0.032588	0.070686	0.443381	0.083851	0.298478	0.186125	0.221560	0.293837	0.644508
8000	-10420.174	1.058	0.129152	0.263612	0.036846	0.061359	0.424323	0.084708	0.284539	0.192084	0.216456	0.306921	0.691606
9000	-10419.701	0.980	0.101173	0.266573	0.035445	0.072158	0.438826	0.085825	0.285541	0.188378	0.229610	0.296471	0.687021
10000	-10423.917	1.015	0.100312	0.289851	0.045985	0.059364	0.422372	0.082115	0.285505	0.176257	0.228230	0.310007	0.684473
11000	-10418.487	0.945	0.107911	0.270677	0.049322	0.063833	0.421602	0.086655	0.279829	0.188085	0.233921	0.298165	0.860128
12000	-10420.169	0.893	0.115085	0.270950	0.038203	0.070506	0.417478	0.087778	0.288131	0.191473	0.231758	0.288638	0.723312
13000	-10419.081	0.922	0.115323	0.269076	0.036184	0.069919	0.429555	0.079943	0.294340	0.187665	0.227043	0.290952	0.784700
14000	-10423.817	1.030	0.112545	0.254842	0.042601	0.077867	0.436797	0.075348	0.283706	0.189549	0.224014	0.302731	0.615981
15000	-10424.879	0.944	0.131641	0.260134	0.043160	0.069779	0.421550	0.073736	0.296187	0.175620	0.219147	0.309046	0.797970
16000	-10426.143	0.940	0.117469	0.266011	0.056463	0.049593	0.441326	0.069139	0.282578	0.203117	0.231372	0.282933	0.792757
17000	-10421.133	0.978	0.134024	0.277374	0.040419	0.056384	0.416233	0.075565	0.289061	0.187968	0.225825	0.297145	0.767063
18000	-10418.290	0.930	0.104450	0.251683	0.041434	0.063649	0.455528	0.083256	0.287086	0.189510	0.226700	0.296704	0.767072
19000	-10420.052	0.972	0.121227	0.274901	0.037023	0.083743	0.414224	0.068881	0.289061	0.187968	0.225825	0.297145	0.758345
20000	-10425.127	0.955	0.099741	0.277386	0.043745	0.069447	0.433059	0.076622	0.292229	0.197483	0.212827	0.297461	0.645034
21000	-10421.087	0.939	0.105737	0.258514	0.039941	0.094773	0.429045	0.071991	0.292778	0.192129	0.217655	0.297438	0.692877
22000	-10421.805	0.926	0.111237	0.293260	0.047595	0.061320	0.409044	0.077544	0.286897	0.197795	0.222410	0.292899	0.797696
23000	-10422.326	0.943	0.123590	0.240213	0.047236	0.048864	0.453312	0.086786	0.291024	0.187438	0.225934	0.295603	0.851381
24000	-10417.974	0.938	0.123674	0.274369	0.051414	0.065387	0.413009	0.072146	0.291024	0.187438	0.225934	0.295603	0.801620
25000	-10422.454	0.996	0.132415	0.249036	0.036744	0.063052	0.457012	0.061741	0.299053	0.171847	0.226435	0.302665	0.607659
26000	-10424.506	0.892	0.122118	0.235061	0.042240	0.063788	0.462004	0.074790	0.302331	0.170502	0.220011	0.307156	0.812245
27000	-10420.001	0.953	0.128264	0.263415	0.040470	0.058989	0.432138	0.076724	0.279181	0.190422	0.234369	0.296028	0.824956

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

We can query the joint distribution marginally with respect to any parameter.

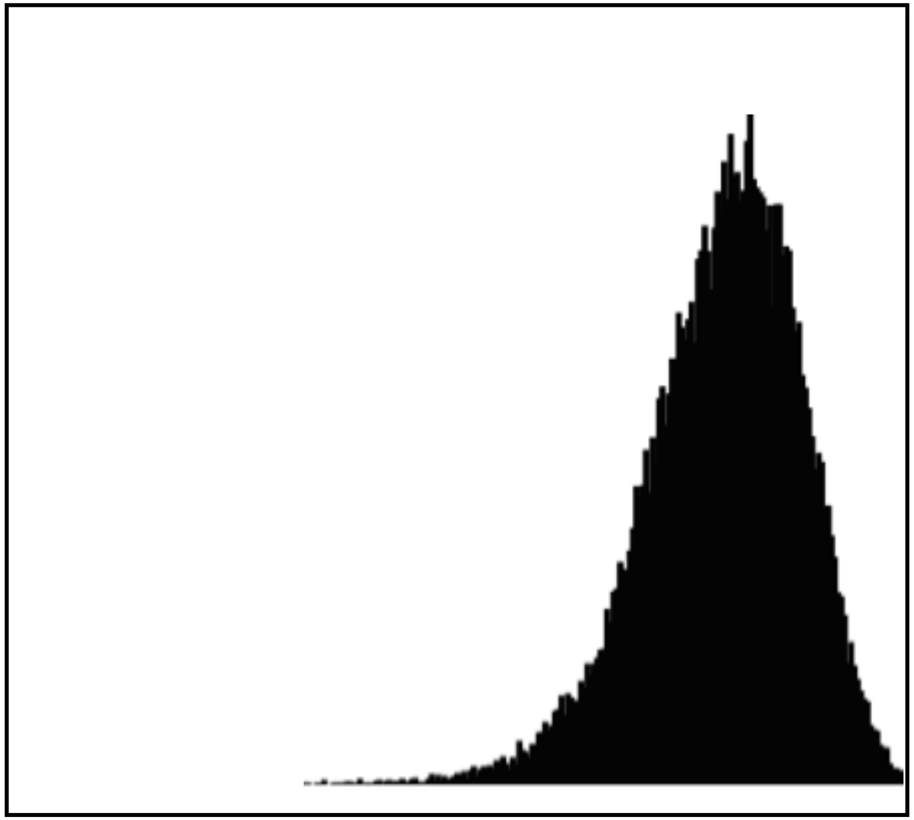
[ID: 2325481386]

Gen	LnL	TL	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)	alpha
1	-13413.769	1.313	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.250000	0.250000	0.250000	0.250000	0.500000
1000	-10429.772	0.904	0.100364	0.271178	0.057126	0.095681	0.404818	0.070833	0.276201	0.173231	0.228359	0.322209	0.845634
2000	-10420.654	0.980	0.115937	0.254216	0.041309	0.051039	0.455344	0.082157	0.291050	0.181003	0.231042	0.296904	0.670406
3000	-10417.930	0.961	0.137253	0.264348	0.037891	0.056962	0.426295	0.077251	0.291050	0.181003	0.231042	0.296904	0.901480
4000	-10423.816	0.925	0.101065	0.273786	0.035266	0.067623	0.441301	0.080958	0.290603	0.185952	0.231800	0.291644	0.859284
5000	-10425.264	1.002	0.135985	0.259584	0.048509	0.057733	0.430436	0.067753	0.289106	0.189615	0.210373	0.310906	0.671675
6000	-10421.366	0.962	0.119016	0.268203	0.041284	0.062913	0.415543	0.093041	0.281133	0.187367	0.234148	0.297353	0.824395
7000	-10417.840	0.981	0.123308	0.246185	0.032588	0.070686	0.443381	0.083851	0.298478	0.186125	0.221560	0.293837	0.644508
8000	-10420.174	1.058	0.129152	0.263612	0.036846	0.061359	0.424323	0.084708	0.284539	0.192084	0.216456	0.306921	0.691606
9000	-10419.701	0.980	0.101173	0.266573	0.035445	0.072158	0.438826	0.085825	0.285541	0.188378	0.229610	0.296471	0.687021
10000	-10423.917	1.015	0.100312	0.289851	0.045985	0.059364	0.422372	0.082115	0.285505	0.176257	0.228230	0.310007	0.684473
11000	-10418.487	0.945	0.107911	0.270677	0.049322	0.063833	0.421602	0.086655	0.279829	0.188085	0.233921	0.298165	0.860128
12000	-10420.169	0.893	0.115085	0.270950	0.038203	0.070506	0.417478	0.087778	0.288131	0.191473	0.231758	0.288638	0.723312
13000	-10419.081	0.922	0.115323	0.269076	0.036184	0.069919	0.429555	0.079943	0.294340	0.187665	0.227043	0.290952	0.784700
14000	-10423.817	1.030	0.112545	0.254842	0.042601	0.077867	0.436797	0.075348	0.283706	0.189549	0.224014	0.302731	0.615981
15000	-10424.879	0.944	0.131641	0.260134	0.043160	0.069779	0.421550	0.073736	0.296187	0.175620	0.219147	0.309046	0.797970
16000	-10426.143	0.940	0.117469	0.266011	0.056463	0.049593	0.441326	0.069139	0.282578	0.203117	0.231372	0.282933	0.792757
17000	-10421.133	0.978	0.134024	0.277374	0.040419	0.056384	0.416233	0.075565	0.289061	0.187968	0.225825	0.297145	0.767063
18000	-10418.290	0.930	0.104450	0.251683	0.041434	0.063649	0.455528	0.083256	0.287086	0.189510	0.226700	0.296704	0.767072
19000	-10420.052	0.972	0.121227	0.274901	0.037023	0.083743	0.414224	0.068881	0.289061	0.187968	0.225825	0.297145	0.758345
20000	-10425.127	0.955	0.099741	0.277386	0.043745	0.069447	0.433059	0.076622	0.292229	0.197483	0.212827	0.297461	0.645034
21000	-10421.087	0.939	0.105737	0.258514	0.039941	0.094773	0.429045	0.071991	0.292778	0.192129	0.217655	0.297438	0.692877
22000	-10421.805	0.926	0.111237	0.293260	0.047595	0.061320	0.409044	0.077544	0.286897	0.197795	0.222410	0.292899	0.797696
23000	-10422.326	0.943	0.123590	0.240213	0.047236	0.048864	0.453312	0.086786	0.291024	0.187438	0.225934	0.295603	0.851381
24000	-10417.974	0.938	0.123674	0.274369	0.051414	0.065387	0.413009	0.072146	0.291024	0.187438	0.225934	0.295603	0.801620
25000	-10422.454	0.996	0.132415	0.249036	0.036744	0.063052	0.457012	0.061741	0.299053	0.171847	0.226435	0.302665	0.607659
26000	-10424.506	0.892	0.122118	0.235061	0.042240	0.063788	0.462004	0.074790	0.302331	0.170502	0.220011	0.307156	0.812245
27000	-10420.001	0.953	0.128264	0.263415	0.040470	0.058989	0.432138	0.076724	0.279181	0.190422	0.234369	0.296028	0.824956

Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

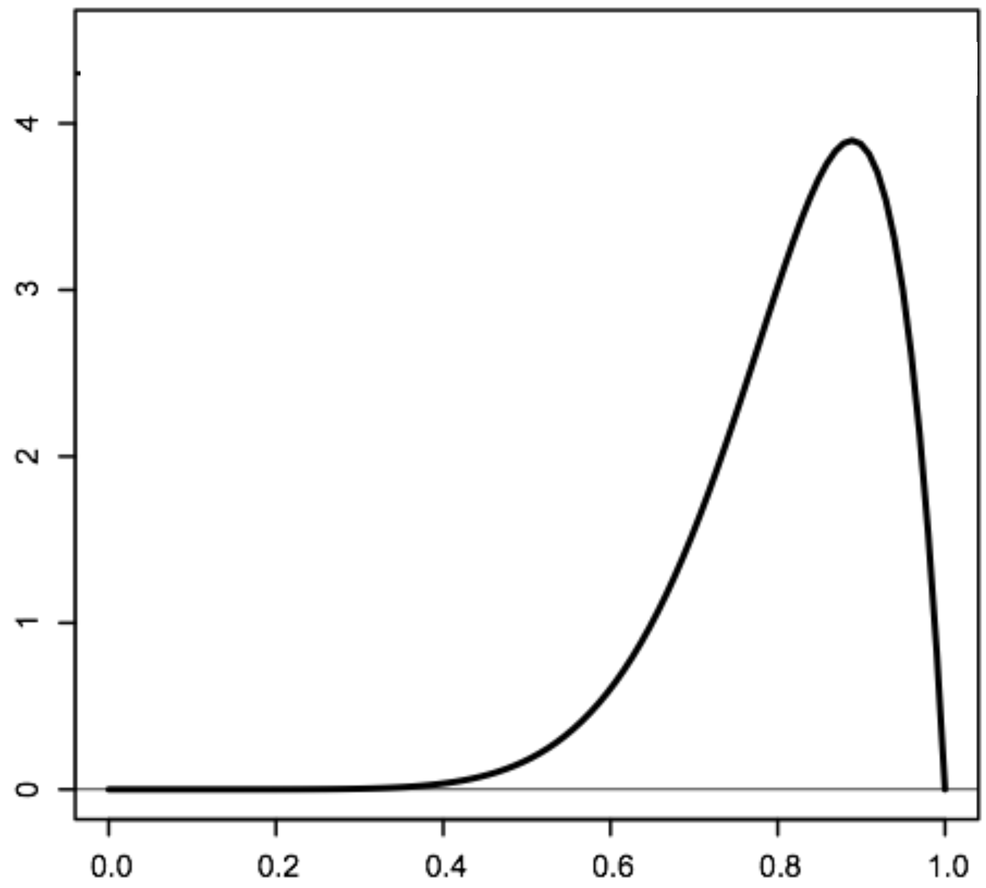
We can do this by simply constructed a histogram for any column in the file
this provides an estimate of its marginal posterior probability density



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

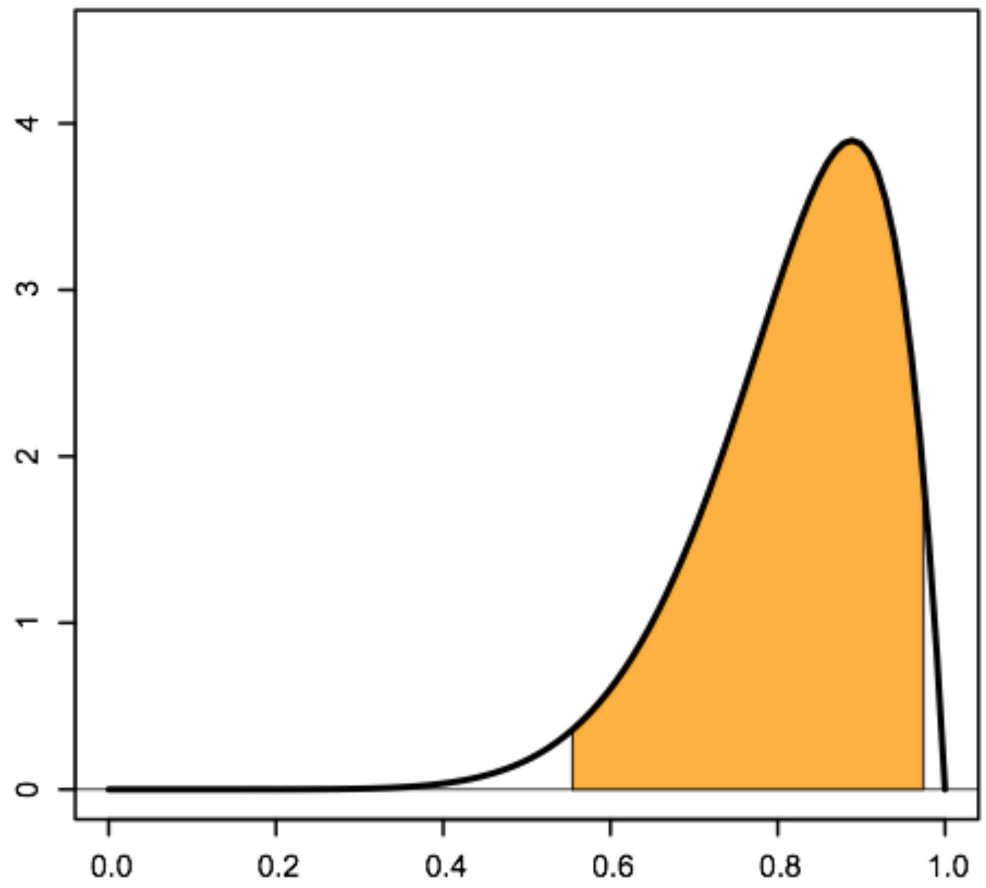
We can do this by simply constructed a histogram for any column in the file this provides an estimate of its marginal posterior probability density



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

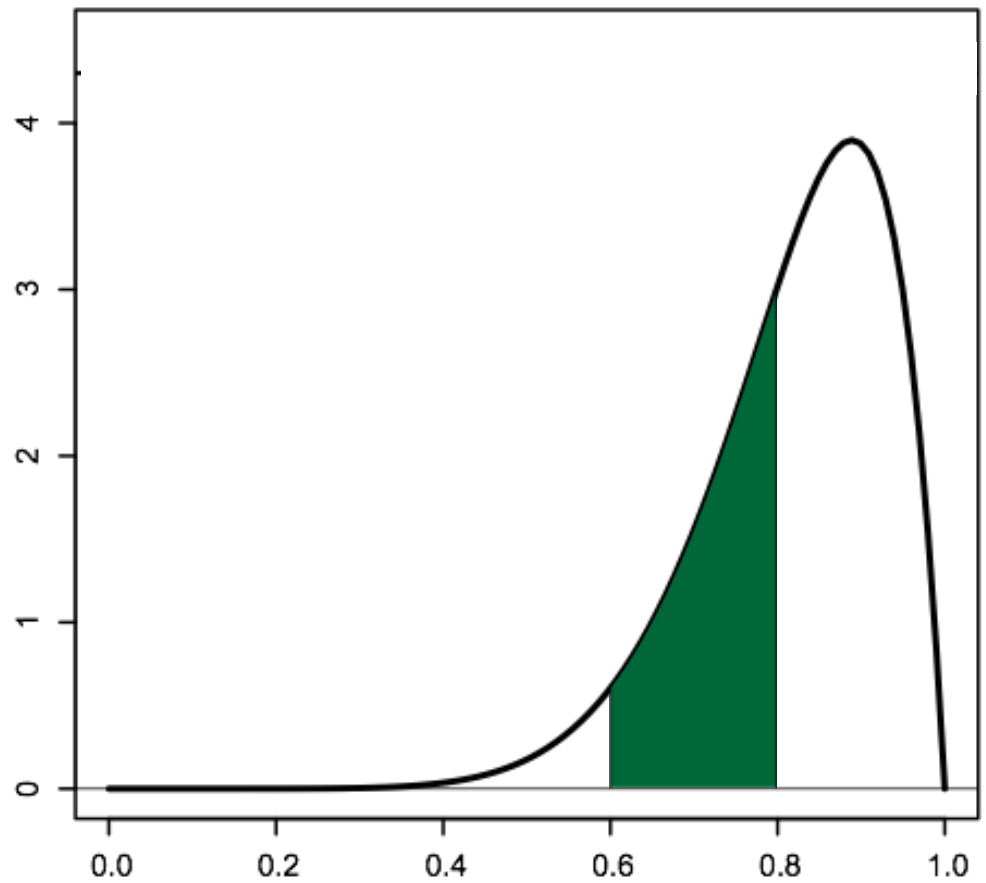
We can easily summarize aspects of the marginal posterior probability density: *e.g.*, to summarize the 95% credible interval.



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

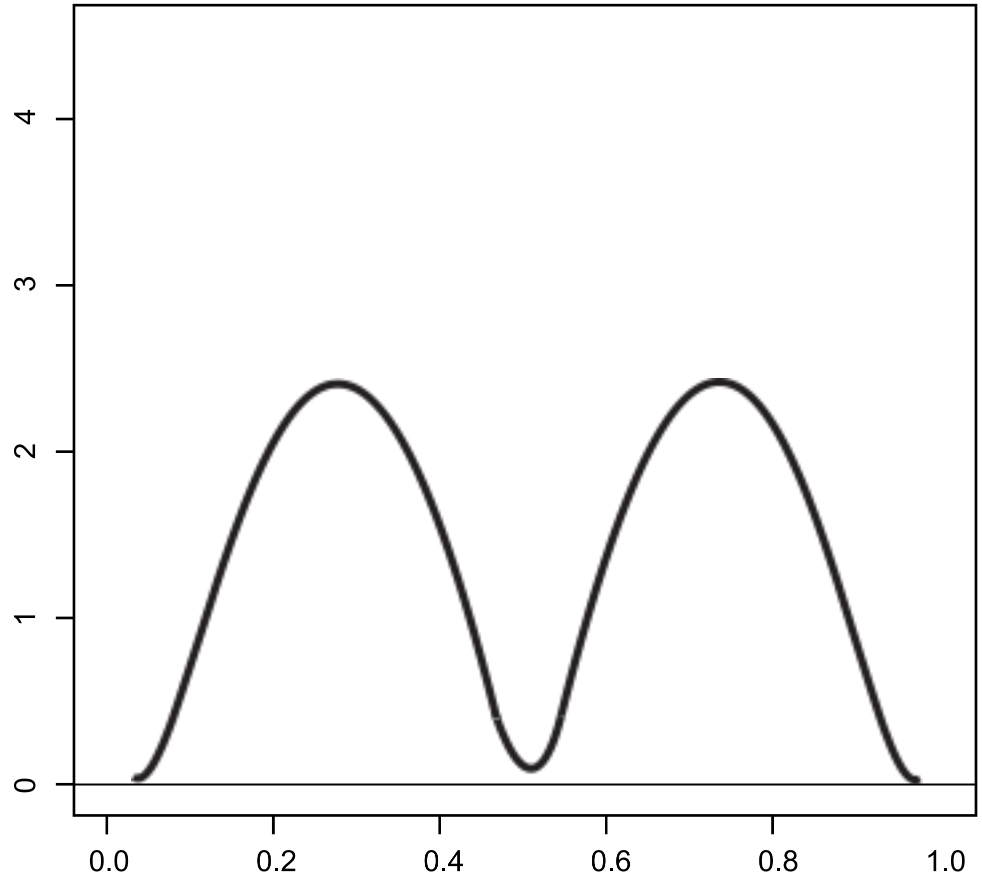
We can easily summarize aspects of the marginal posterior probability density: *e.g.*, or the probability within some arbitrary interval of interest (0.6–0.8).



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

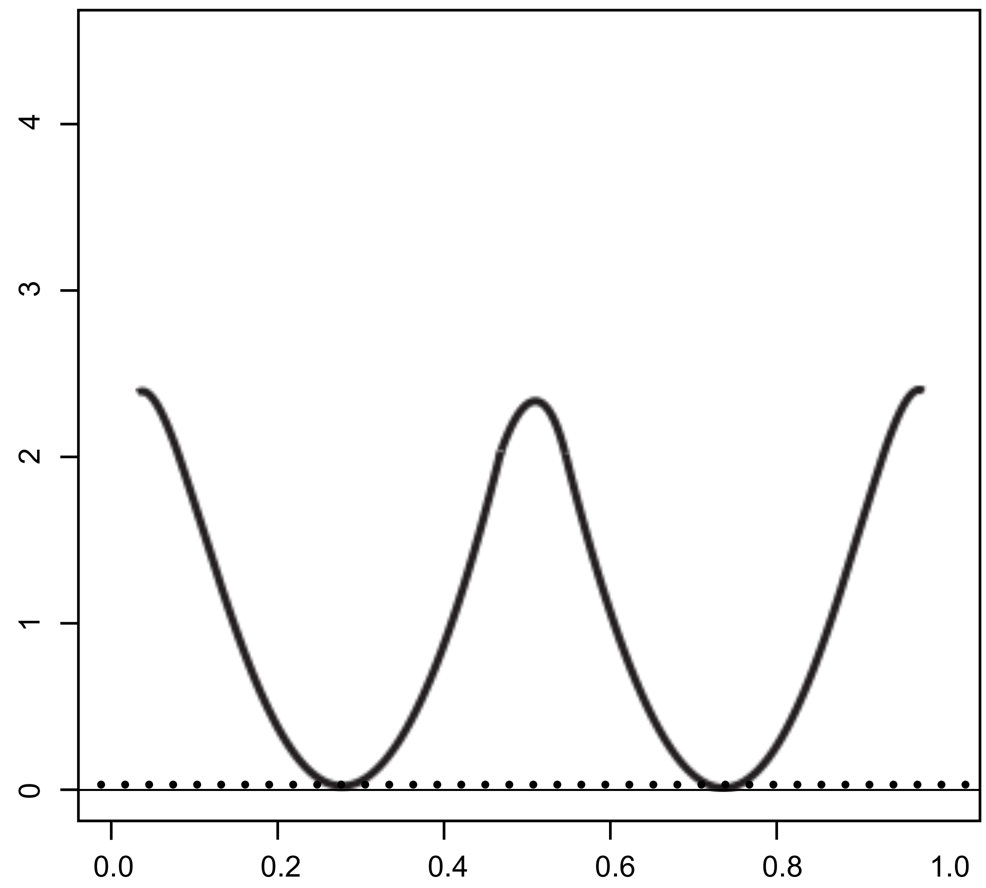
We can easily summarize aspects of the marginal posterior probability density: *e.g.*, or we can summarize the highest posterior density (HPD) interval.



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

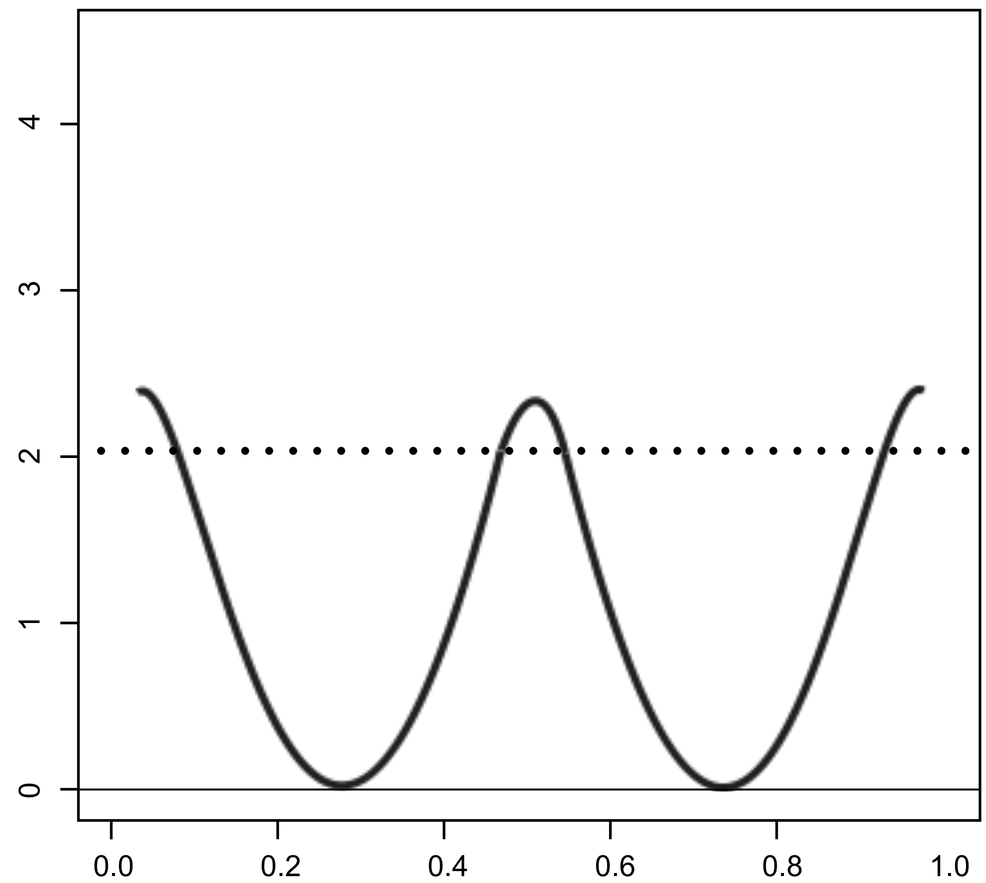
We can easily summarize aspects of the marginal posterior probability density: *e.g.*, or we can summarize the highest posterior density (HPD) interval.



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

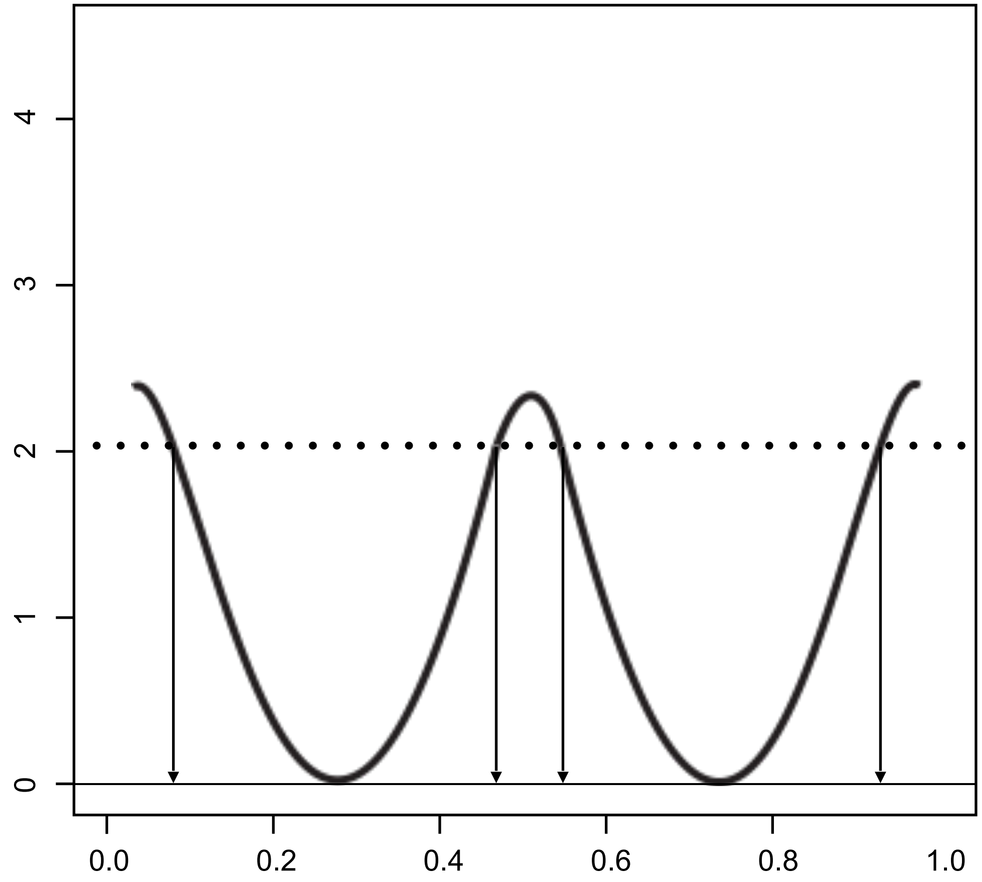
We can easily summarize aspects of the marginal posterior probability density: *e.g.*, or we can summarize the highest posterior density (HPD) interval.



Approximating the Joint Posterior Probability Density using MCMC

Samples from the MCMC simulation approximate the joint posterior

We can easily summarize aspects of the marginal posterior probability density: *e.g.*, or we can summarize the highest posterior density (HPD) interval.



Outline

I. Introduction to Bayesian inference

What is Bayesian inference?

- Deriving Bayes theorem
- Two non-phylogenetic examples
- Bayesian inference of phylogeny

II. The Bayesian hard sell

What's the deal with priors?

Learning to embrace your inner Bayesian

III. Numerical algorithms for Bayesian inference

Markov-chain Monte Carlo (MCMC)

- Metropolis-Hastings algorithm
- Metropolis-Coupled algorithm

Summarizing posterior samples

Outline

IV. Diagnosing MCMC performance

Motivation and overview of the basics

V. MCMC Diagnostics

General strategies:

- diagnostics based on single chains
- diagnostics based on multiple, replicate chains

Approximating the Joint Posterior Probability Density using MCMC

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

Approximating the Joint Posterior Probability Density using MCMC

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 particular case

Approximating the Joint Posterior Probability Density using MCMC

MCMC in theory and practice

MCMC in theory...

an **<u>appropriately constructed</u>** and **<u>adequately run</u>** 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 particular case

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

A.

NEVER!

Approximating the Joint Posterior Probability Density using MCMC

MCMC performance

It is not sufficient to merely be deeply concerned about MCMC performance...
you need to be completely obsessed about it!

Approximating the Joint Posterior Probability Density using MCMC

MCMC performance

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

Approximating the Joint Posterior Probability Density using MCMC

MCMC performance

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

careless **careful** **paranoid**

Approximating the Joint Posterior Probability Density using MCMC

Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review

Mary Kathryn COWLES and Bradley P. CARLIN

A critical issue for users of Markov chain Monte Carlo (MCMC) methods in applications is how to determine when it is safe to stop sampling and use the samples to estimate characteristics of the distribution of interest. Research into methods of computing theoretical convergence bounds holds promise for the future but to date has yielded relatively little of practical use in applied work. Consequently, most MCMC users address the convergence problem by applying diagnostic tools to the output produced by running their samplers. After giving a brief overview of the area, we provide an expository review of 13 convergence diagnostics, describing the theoretical basis and practical implementation of each. We then compare their performance in two simple models and conclude that all of the methods can fail to detect the sorts of convergence failure that they were designed to identify. We thus recommend a combination of strategies aimed at evaluating and accelerating MCMC sampler convergence, including applying diagnostic procedures to a small number of parallel chains, monitoring autocorrelations and cross-correlations, and modifying parameterizations or sampling algorithms appropriately. We emphasize, however, that it is not possible to say with certainty that a finite sample from an MCMC algorithm is representative of an underlying stationary distribution.

KEY WORDS: Autocorrelation; Gibbs sampler; Metropolis-Hastings algorithm.

J. Am. Stat. Soc. (1996)

Approximating the Joint Posterior Probability Density using MCMC

Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review

Mary Kathryn COWLES and Bradley P. CARLIN

A critical issue for users of Markov chain Monte Carlo (MCMC) methods in applications is how to determine when it is safe to stop sampling and use the samples to estimate characteristics of the distribution of interest. Research into methods of computing theoretical convergence bounds holds promise for the future but to date has yielded relatively little of practical use in applied work. Consequently, most MCMC users address the convergence problem by applying diagnostic tools to the output produced by

…under simulation, all MCMC diagnostics may fail to detect the exact problems that they were specifically designed to identify…

…therefore, it is critical to use a combination of tools to detect MCMC failure

a finite sample from an MCMC algorithm is representative of an underlying stationary distribution.

KEY WORDS: Autocorrelation; Gibbs sampler; Metropolis-Hastings algorithm.

J. Am. Stat. Soc. (1996)

Outline

IV. Diagnosing MCMC performance

Motivation and overview of the basics

V. MCMC Diagnostics

General strategies:

- diagnostics based on single chains
- diagnostics based on multiple, replicate chains

Outline

IV. Diagnosing MCMC performance

Motivation and overview of the basics

V. MCMC Diagnostics

General strategies:

- diagnostics based on single chains
- diagnostics based on multiple, replicate chains

Assessing MCMC Performance: Two Main Issues

1. Convergence

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

Assessing MCMC Performance: Two Main Issues

1. Convergence

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

2. Mixing

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

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

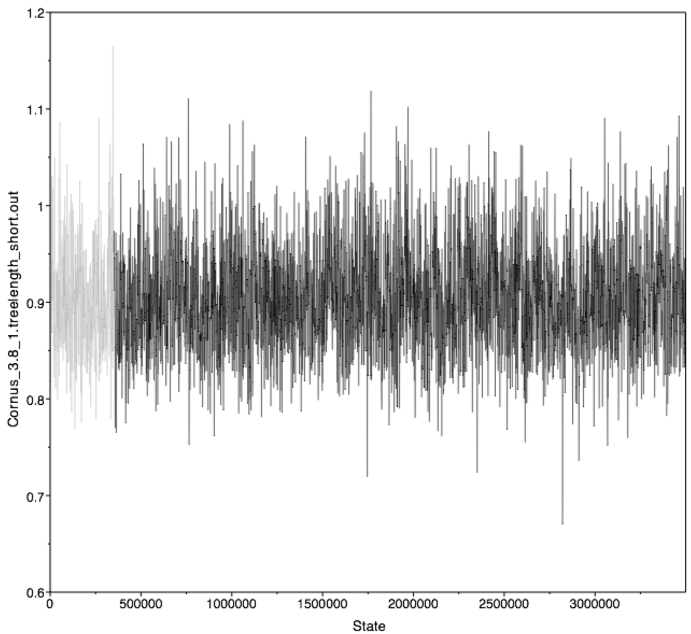
Continuous parameters (*e.g.*, substitution rates)

- some parameters are more reliable than others
- steps may occur!

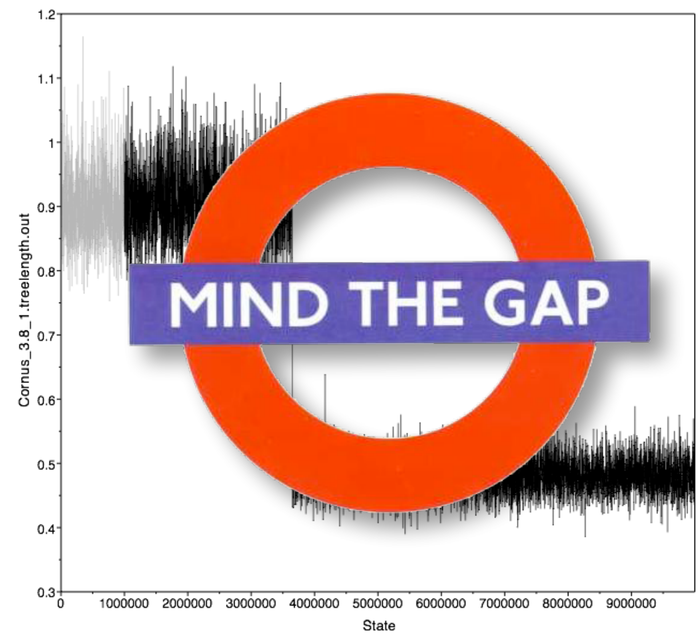
Assessing MCMC Performance: Based on Single Chains

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

all looks good…



until it doesn't



fast* slow*

lnL base freq. sub. rates ASRV TL topology

*somewhat data-set dependent

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

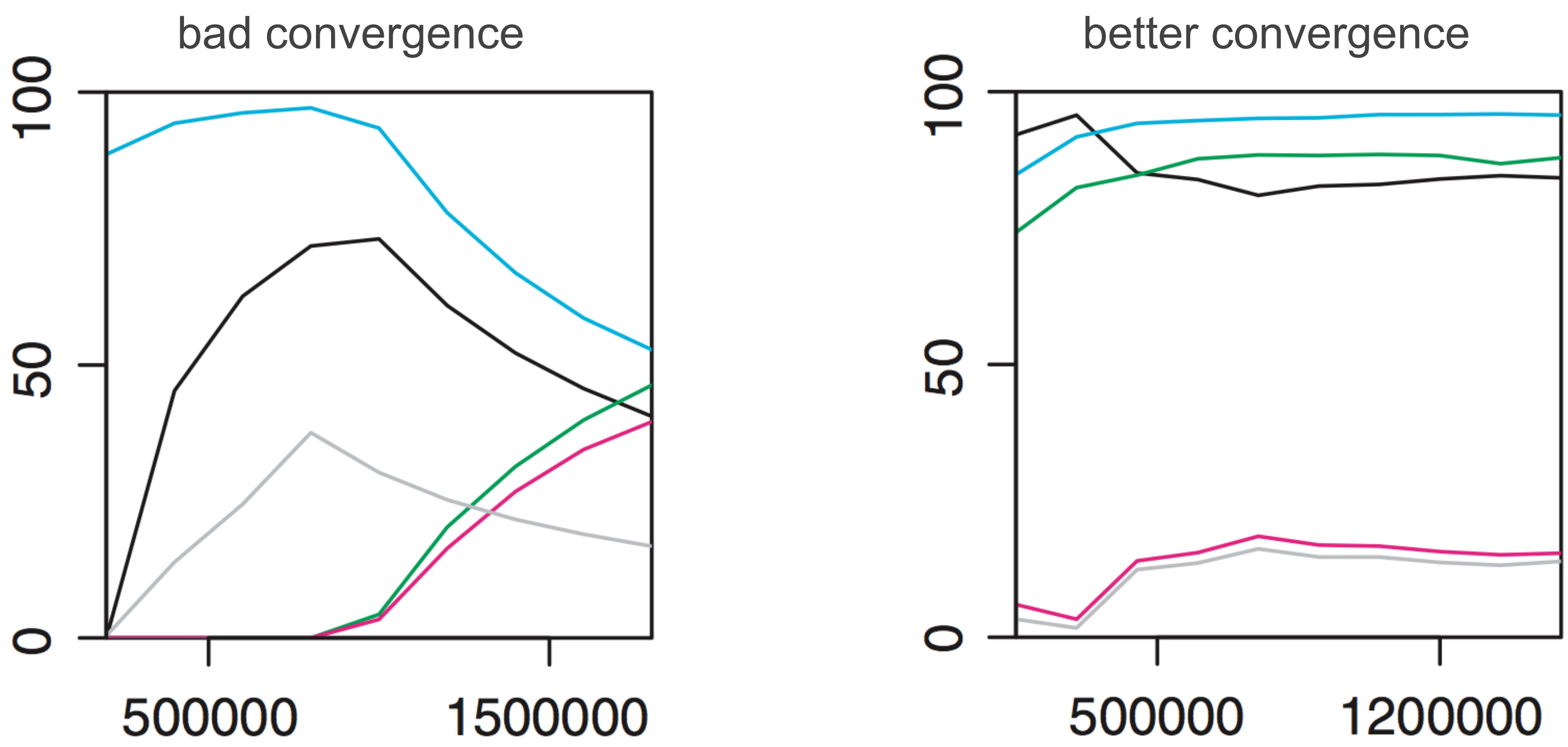
- some parameters are more reliable than others
- steps may occur!

Discrete parameters (*e.g.*, bi-partitions)

- some parameters are more reliable than others
- steps may occur!

Assessing MCMC Performance: Based on Single Chains

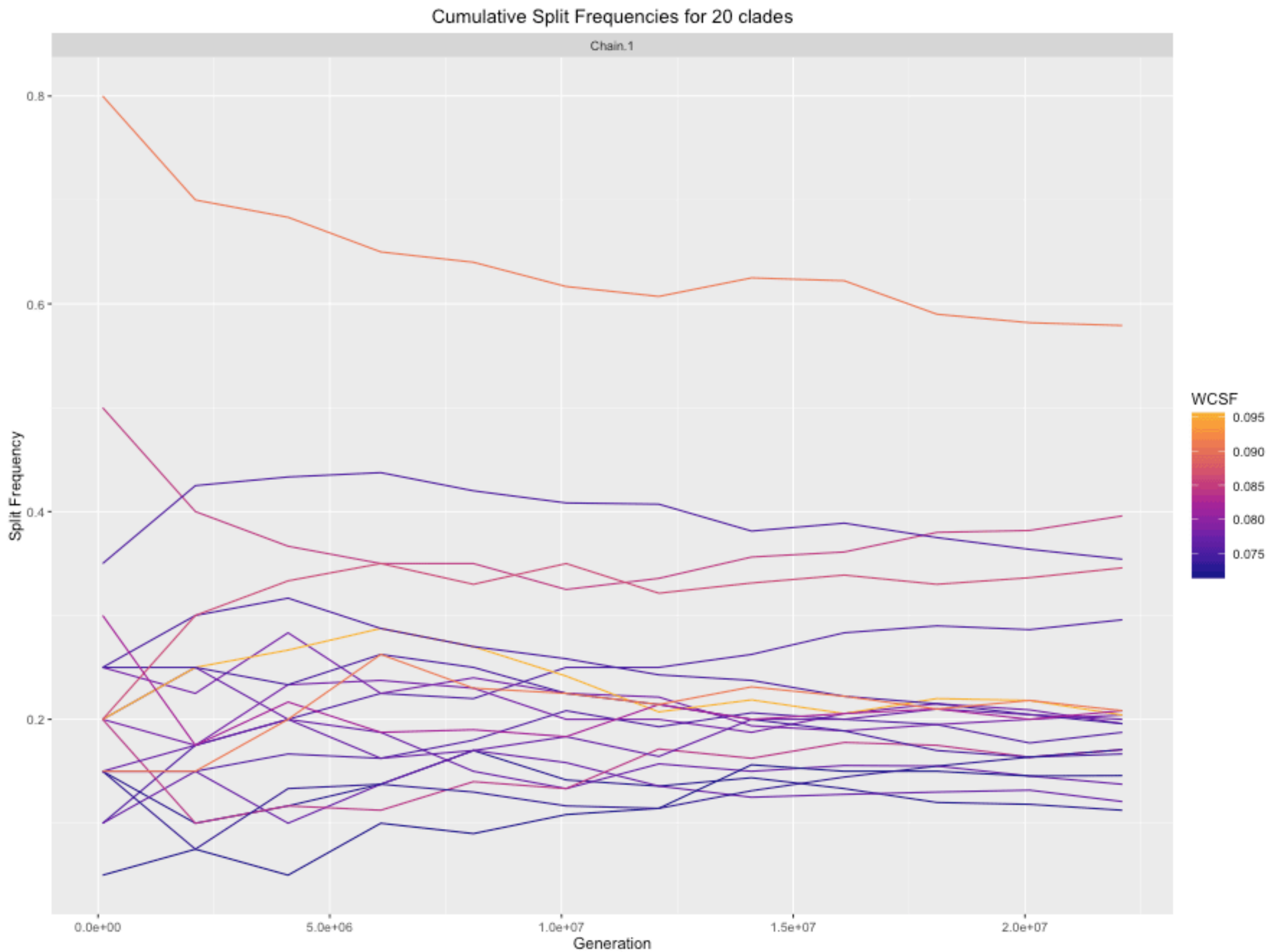
Example: AWTY plots of cumulative bi-partition frequency of 5 nodes



Nylander *et al*. (2008)

Assessing MCMC Performance: Based on Single Chains

Example: **RWTY** plots of cumulative bi-partition frequency of 20 nodes



Warren, Geneva & Lanfear (2017)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke (1992)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters

- A test for equality of the means of the first and last part of a Markov chain (by default the first 10% and the last 50%)

Geweke (1992)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters

- A test for equality of the means of the first and last part of a Markov chain (by default the first 10% and the last 50%)
- If the samples are drawn from the stationary distribution, the two means should equal and Geweke's statistic has an asymptotically standard normal distribution

Geweke (1992)

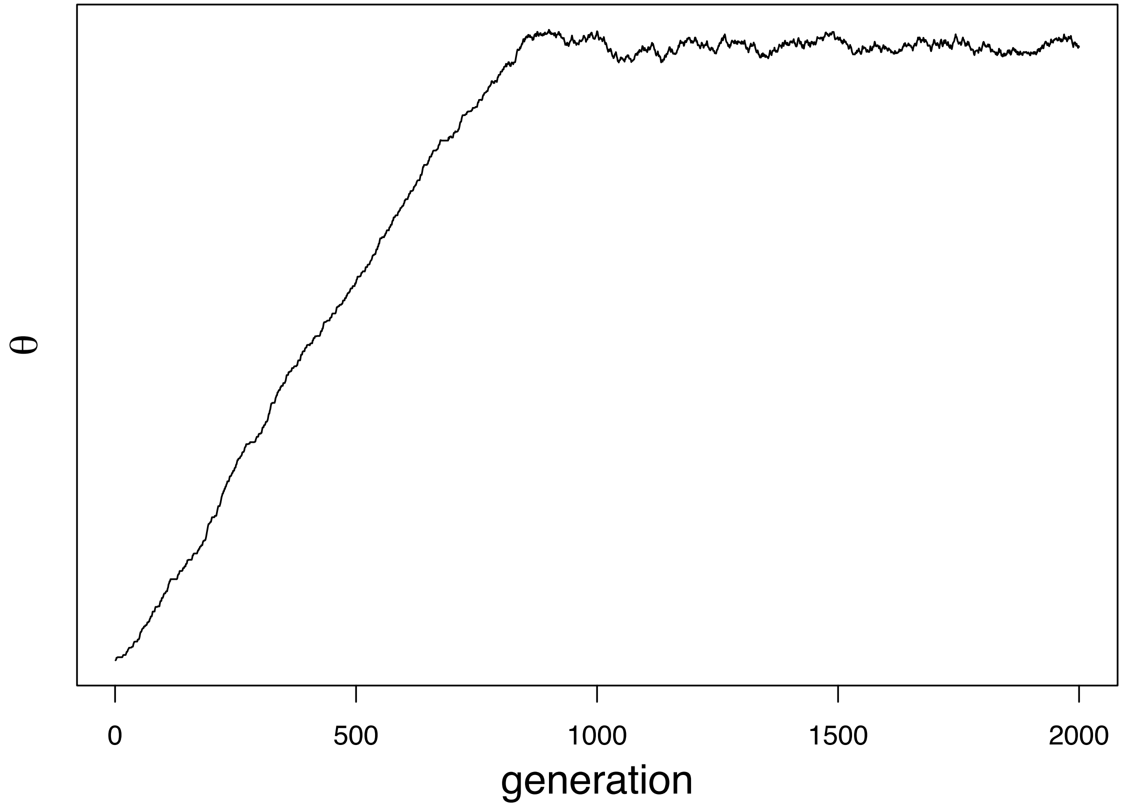
Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters



Geweke (1992)

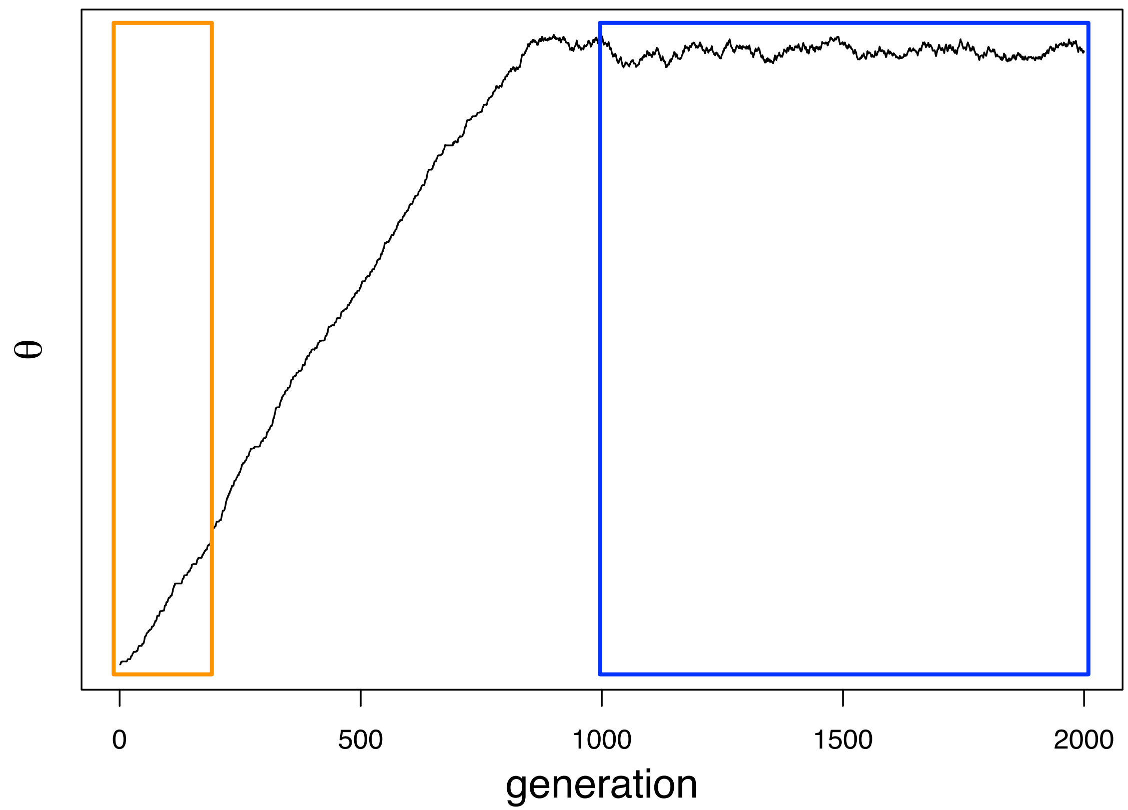
Assessing MCMC Performance: Based on Single Chains

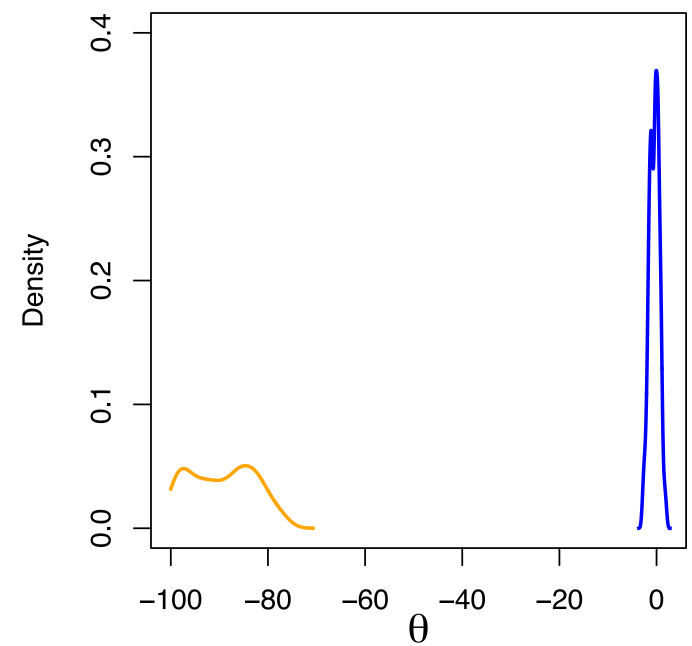
1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters





Geweke (1992)

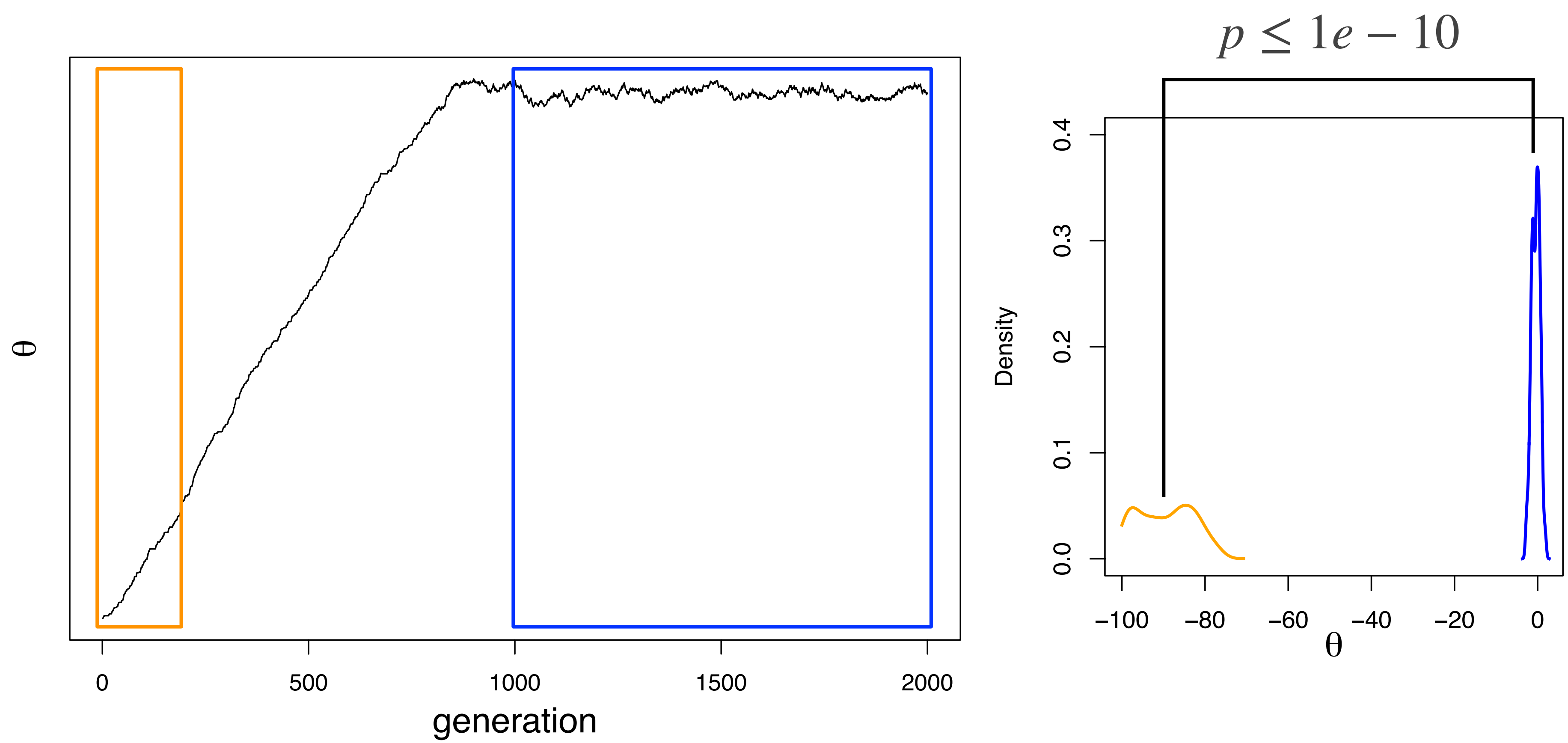
Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters



Geweke (1992)

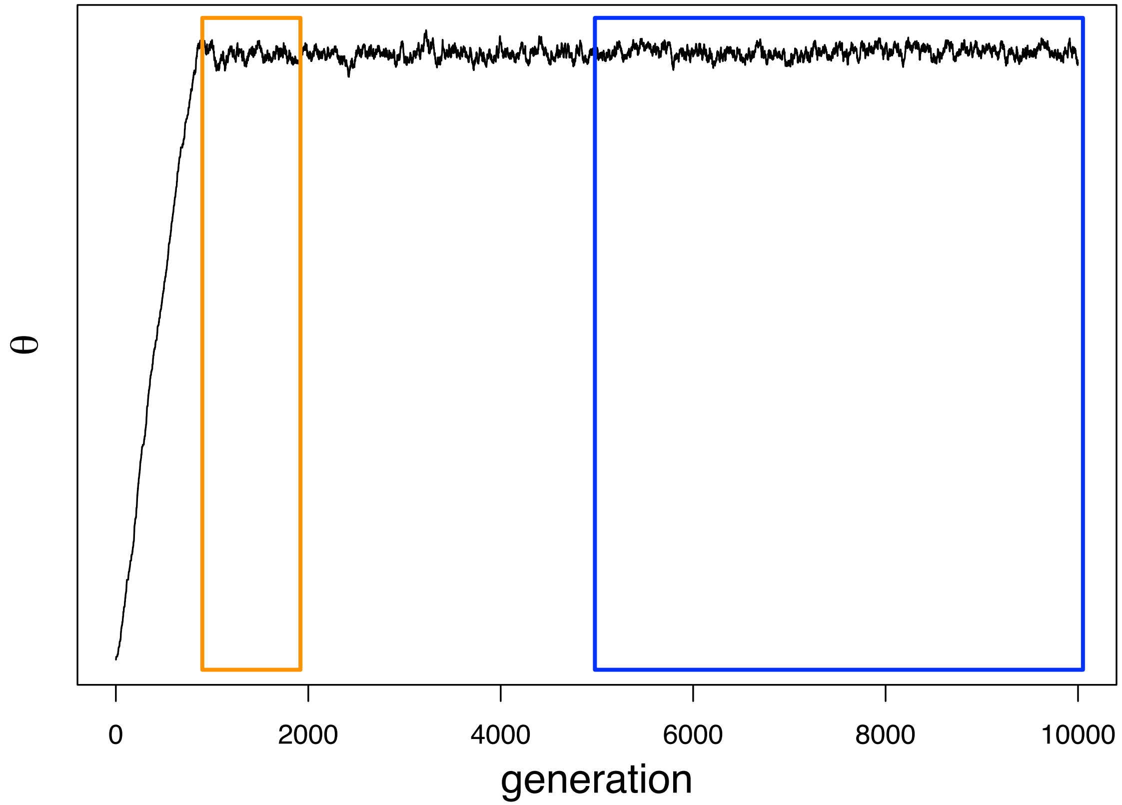
Assessing MCMC Performance: Based on Single Chains

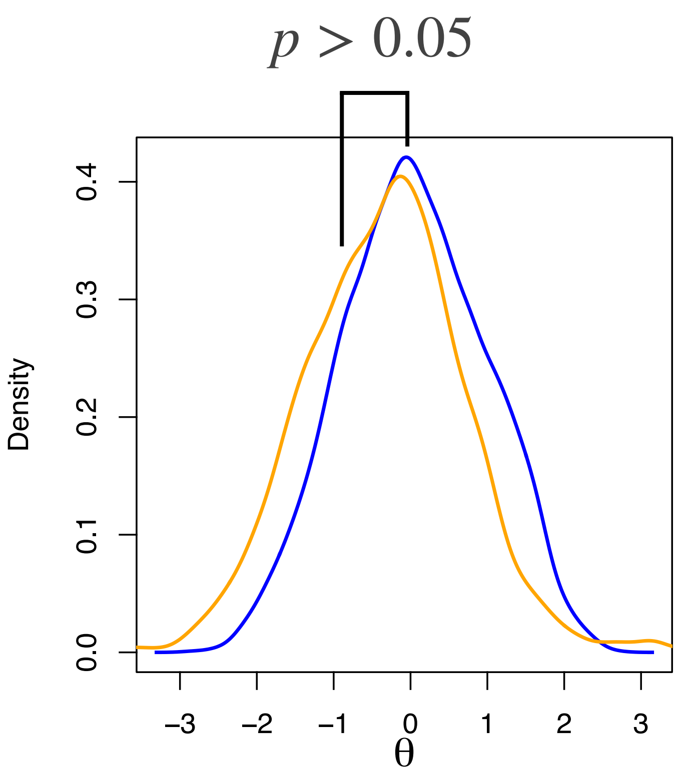
1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Continuous or discrete parameters





Geweke (1992)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Heidelberg-Welch diagnostic (coda)

Continuous or discrete parameters

- Uses the Cramer-von Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution

Heidelberg & Welch (1983)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Heidelberg-Welch diagnostic (coda)

Continuous or discrete parameters

- Uses the Cramer-von Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution
- This test is successively applied, first to the whole chain, then after discarding the first 10%, 20%, $\ldots$ of the samples until either the null hypothesis is accepted, or 50% of the chain has been discarded

Heidelberg & Welch (1983)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Heidelberg-Welch diagnostic (coda)

Continuous or discrete parameters

- Uses the Cramer-von Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution
- This test is successively applied, first to the whole chain, then after discarding the first 10%, 20%, $\ldots$ of the samples until either the null hypothesis is accepted, or 50% of the chain has been discarded
- The latter outcome constitutes “failure” of the test and indicates that a longer run is needed

Heidelberg & Welch (1983)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

Time-series plots of parameter estimates

Geweke diagnostic (coda)

Heidelberg-Welch diagnostic (coda)

Continuous or discrete parameters

- Uses the Cramer-von Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution
- This test is successively applied, first to the whole chain, then after discarding the first 10%, 20%, $\ldots$ of the samples until either the null hypothesis is accepted, or 50% of the chain has been discarded
- The latter outcome constitutes "failure" of the test and indicates that a longer run is needed
- Otherwise, the number of iterations to keep and the number to discard (burn-in) are reported

Heidelberg & Welch (1983)

Assessing MCMC Performance: Based on Single Chains

1. Convergence diagnostics

  Time-series plots of parameter estimates

  Geweke diagnostic (coda)

  Heidelberg-Welch diagnostic (coda)

  (many others)

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

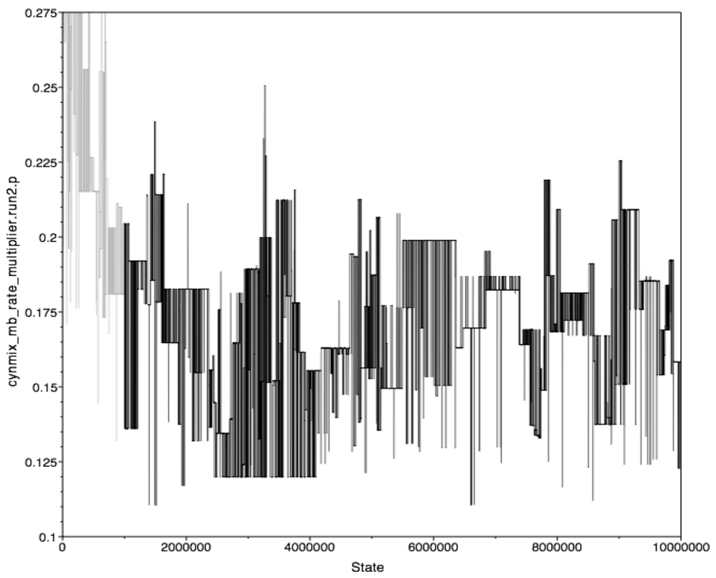
Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

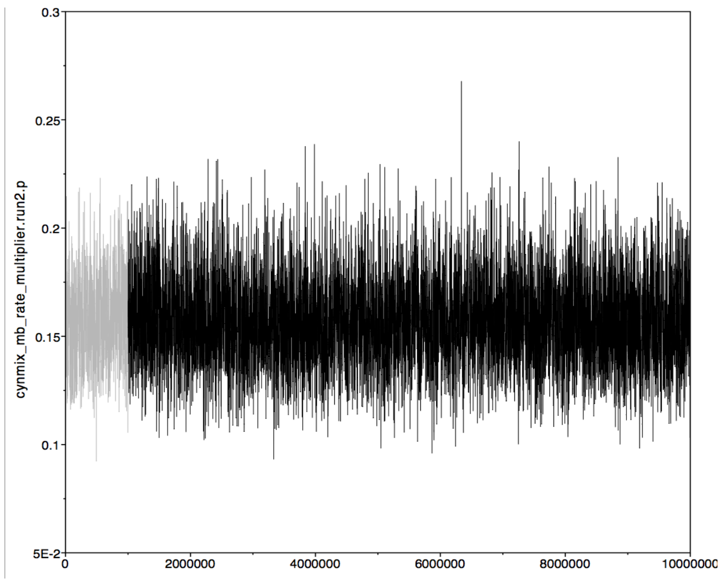
Assessing MCMC Performance: Based on Single Chains

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

bad mixing



better mixing



Assessing MCMC Performance: Based on Single Chains

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

bad mixing

better mixing

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- rates should ideally fall in the ~20–70% range

Assessing MCMC Performance: Based on Single Chains

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: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- 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

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- 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 acceptance rates, decrease scale of tuning parameter (and vice versa)

tuning parameter

proposal weight

prior distribution

parameter

```
# specify a beta prior on x
x ~ dnBeta(1,1)

# place a sliding move on x
moves.append( mvSlide(x, delta = 0.1, weight = 5.0) )
```

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- 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 acceptance rates, decrease scale of tuning parameter (and vice versa)

```
# burn the chain in
mymcmc.burnin(generations = 1000, tuningInterval = 100)
```

adjust tuning parameters

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- rates should ideally fall in the ~20–70% range

Form of the marginal posterior probability densities

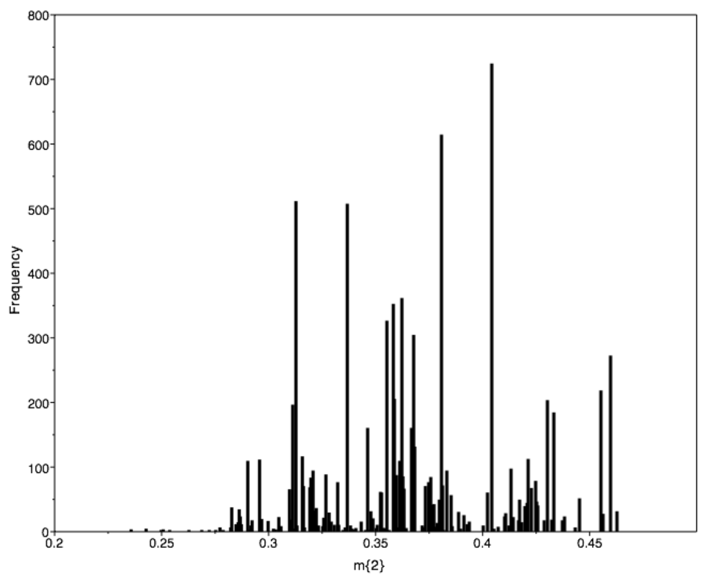
Continuous parameters (*e.g.*, substitution rates)

- beware of porcupine roadkill!

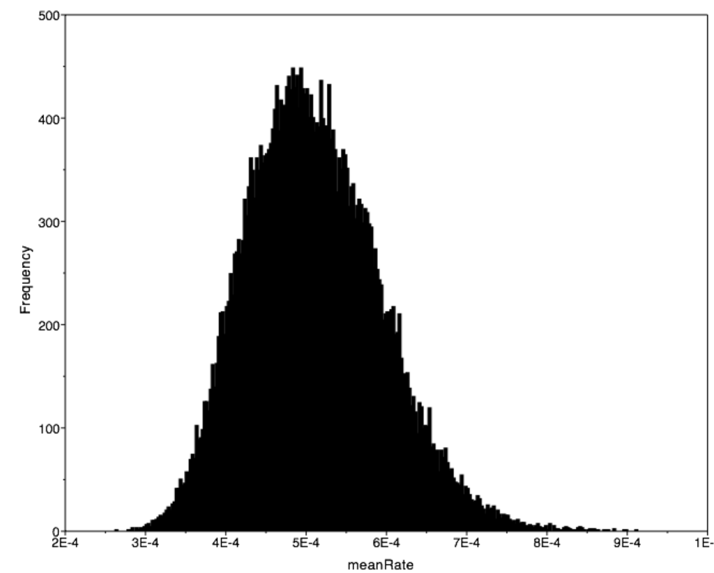
Assessing MCMC Performance: Based on Single Chains

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

bad mixing



better 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
```

```
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
```

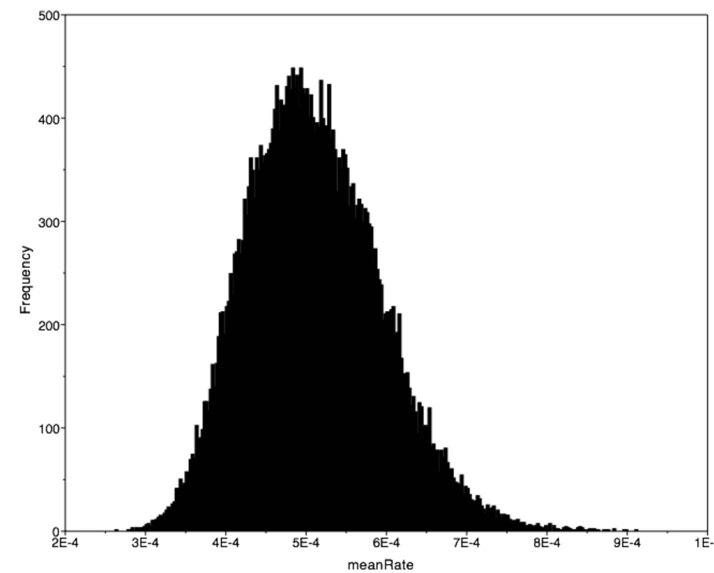
Assessing MCMC Performance: Based on Single Chains

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: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- rates should ideally fall in the ~20–70% range

Form of the marginal posterior probability densities

Continuous parameters (*e.g.*, substitution rates)

- beware of porcupine roadkill!

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g.*, substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- rates should ideally fall in the ~20–70% range

Form of the marginal posterior probability densities

Continuous parameters (*e.g.*, substitution rates)

- beware of porcupine roadkill!

qualitative diagnostics

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Form of the time-series plots of parameter estimates

Continuous parameters (*e.g*., substitution rates)

- warm and fuzzy caterpillars

Acceptance rates of parameter updates

Continuous and discrete parameters (MrBayes, BEAST)

- rates should ideally fall in the ~20–70% range

Form of the marginal posterior probability densities

Continuous parameters (*e.g*., substitution rates)

- beware of porcupine roadkill!

qualitative diagnostics

Autocorrelation time (ACT) of parameter samples

Effective sample size (ACT) of parameter samples

quantitative diagnostics

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Autocorrelation time (ACT) of parameter samples

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Autocorrelation time (ACT) of parameter samples

The lag (number of cycles) it takes for autocorrelation in parameter values to break down.

Assessing MCMC Performance: Based on Single Chains

2. Mixing diagnostics

Autocorrelation time (ACT) of parameter samples

The lag (number of cycles) it takes for autocorrelation in parameter values to break down.

Effective Sample Size (ESS) diagnostic

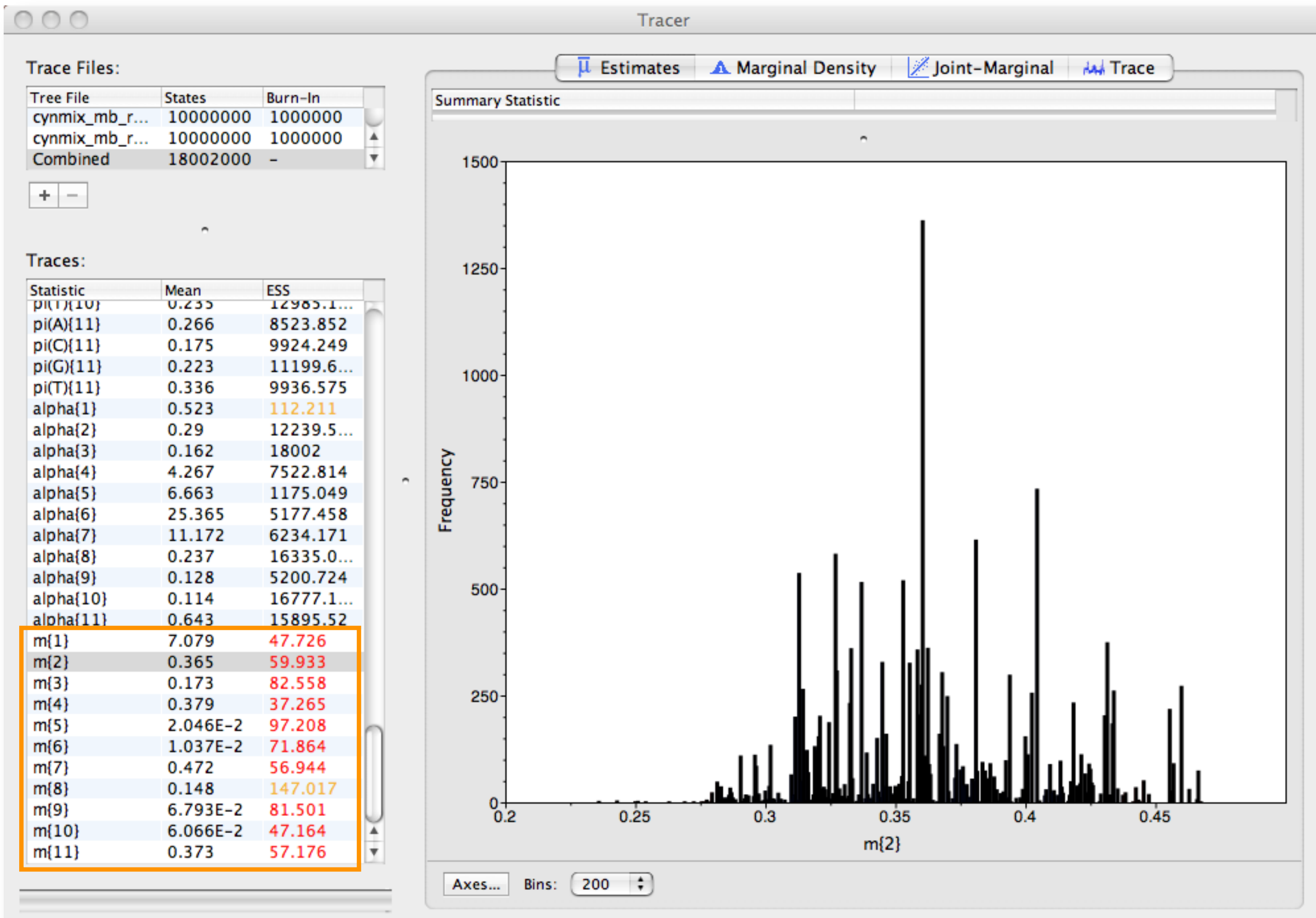
Continuous or discrete parameters

- number of samples/autocorrelation time (ACT)

Assessing MCMC Performance: Based on Single Chains

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

poor mixing



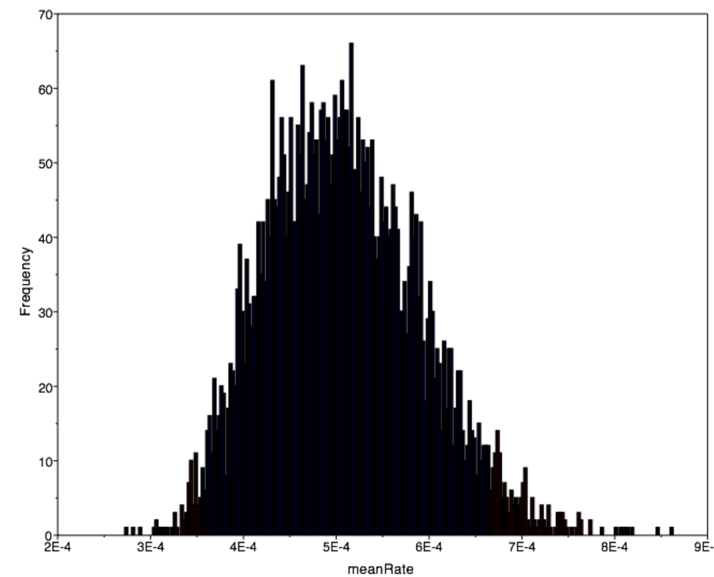
Assessing MCMC Performance: Based on Single Chains

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

poor sampling

1M cycles

better sampling



5M cycles

inadequate chain length/poor mixing

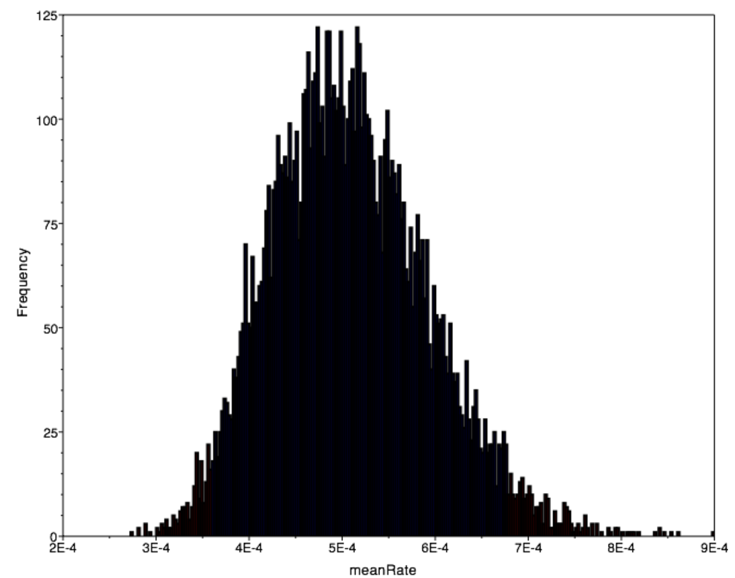
Assessing MCMC Performance: Based on Single Chains

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

poor sampling

1M cycles

better sampling



10M cycles

inadequate chain length/poor mixing

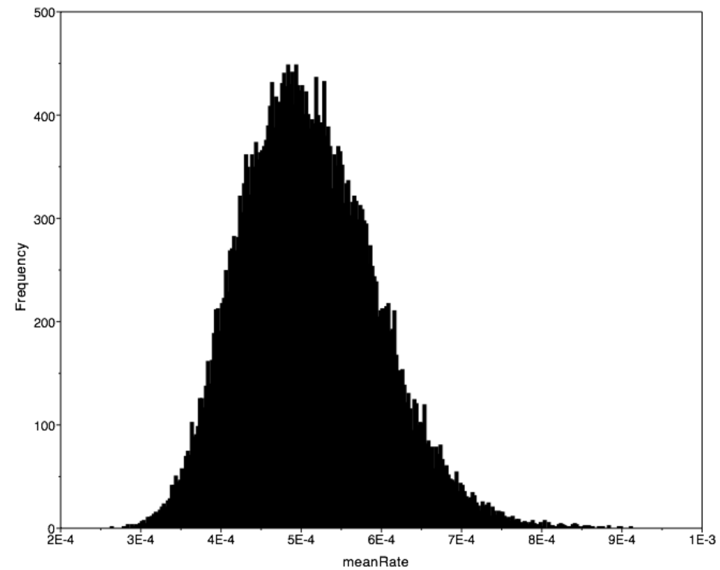
Assessing MCMC Performance: Based on Single Chains

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

poor sampling

1M cycles

better sampling



40M cycles

inadequate chain length/poor mixing

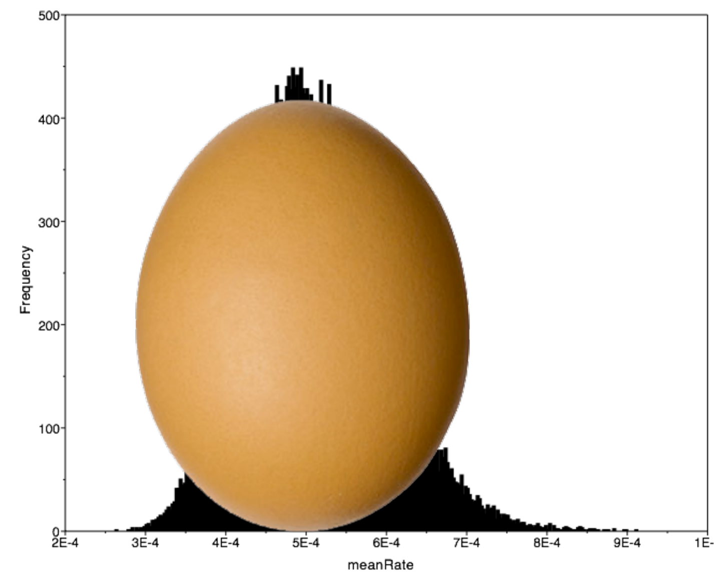
Assessing MCMC Performance: Based on Single Chains

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

poor sampling

1M cycles

better sampling



40M cycles

inadequate chain length/poor mixing

all continuous parameters should be SAE

KDE SAE does not count (use histogram render)

Outline

IV. Diagnosing MCMC performance

Motivation and overview of the basics

V. MCMC Diagnostics

General strategies:

- diagnostics based on single chains
- diagnostics based on multiple, replicate chains

Outline

IV. Diagnosing MCMC performance

Motivation and overview of the basics

V. MCMC Diagnostics

General strategies:

- diagnostics based on single chains
- diagnostics based on multiple, replicate chains

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

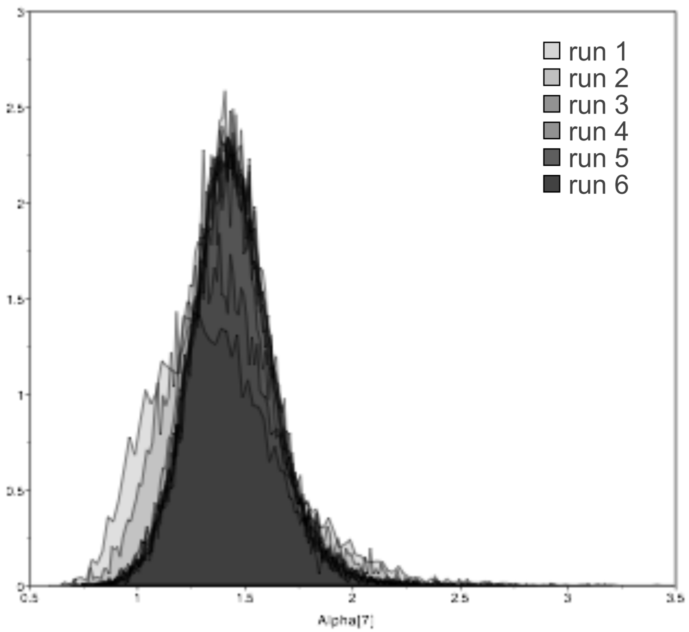
Compare estimates from multiple independent chains

- Form of the marginal posterior densities for all parameters
  - Continuous parameters

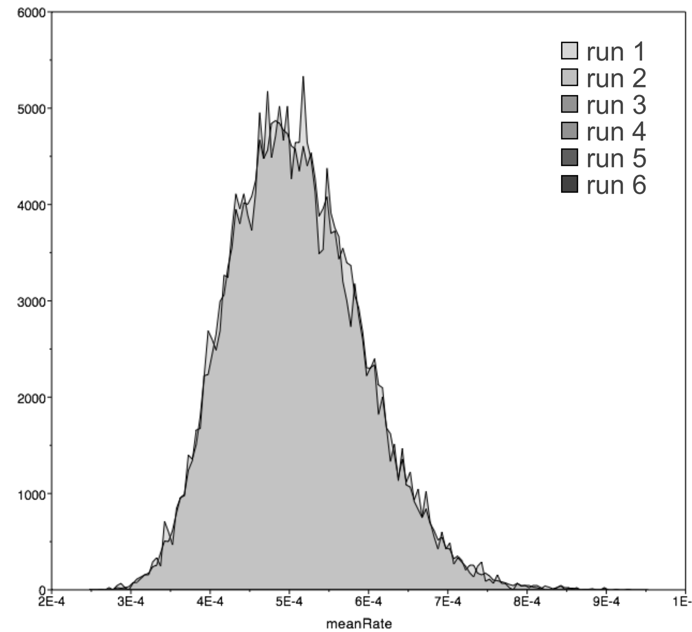
Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: Tracer plots of marginal densities from multiple RevBayes runs

bad convergence



better convergence



Parameter estimates from replicate independent MCMC analyses should be effectively identical.

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Compare estimates from multiple independent chains

- Form of the marginal posterior densities for all parameters
  - Continuous parameters
- PSRF (Gelman–Rubin) diagnostic
  - Continuous and discrete parameters
    1. Run $m \geq 2$ chains of length $2c$ 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 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 Lower	95% Cred. Interval Upper	Median	PSRF *
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 Lower	95% Cred. Interval Upper	Median	PSRF *
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

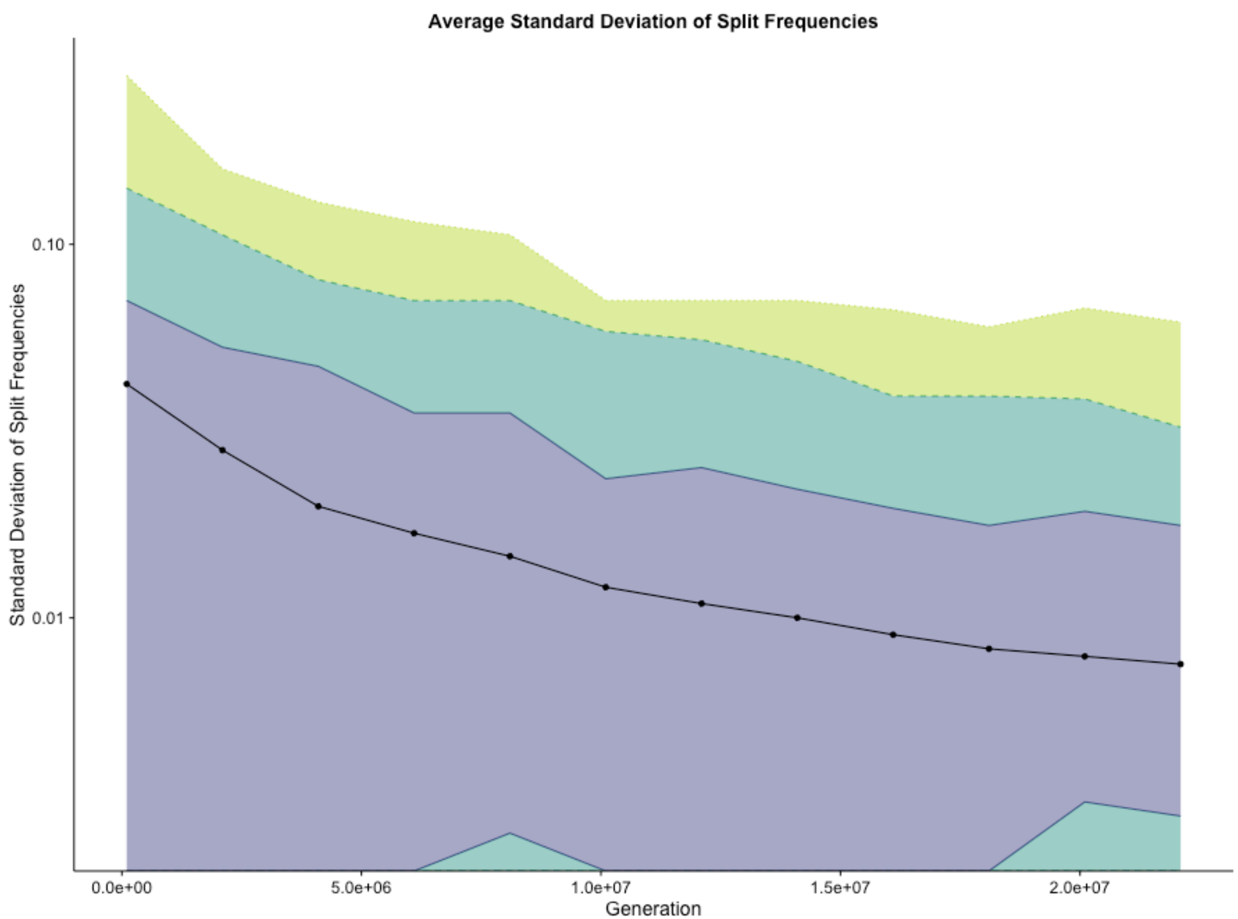
Compare estimates from multiple independent chains

- Form of the marginal posterior densities for all parameters
  - Continuous parameters
- PSRF (Gelman–Rubin) diagnostic
  - Continuous and discrete parameters
- Comparing independent samples of trees
  - ASDSF: similarity of trees sampled by paired, independent chains

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: ASDSF

The overall similarity of the trees sampled by two independent, simultaneous MCMC analyses



Warren, Geneva & Lanfear (2017)

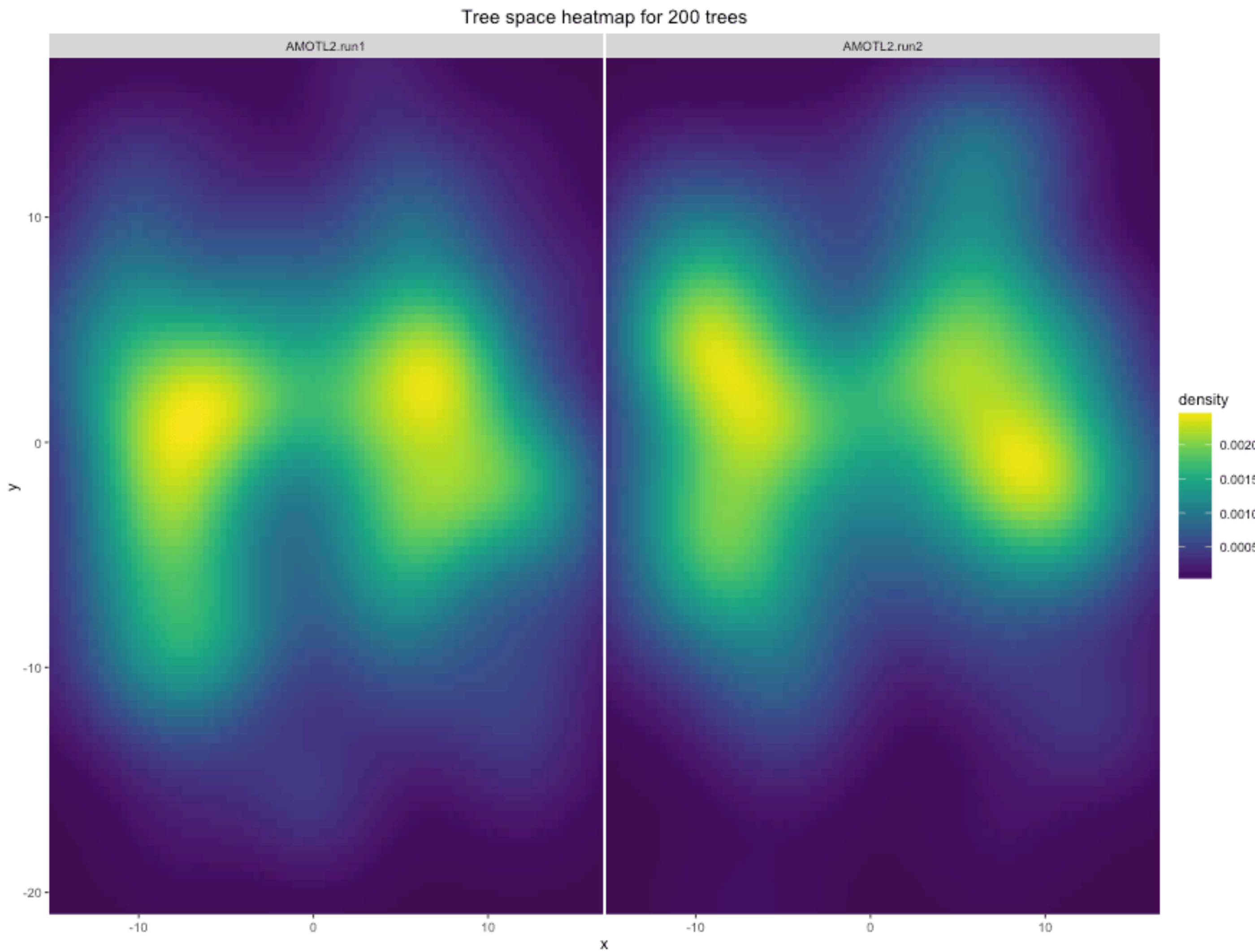
Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Compare estimates from multiple independent chains

- Form of the marginal posterior densities for all parameters
  - Continuous parameters
- PSRF (Gelman–Rubin) diagnostic
  - Continuous and discrete parameters
- Comparing independent samples of trees
  - ASDSF: similarity of trees sampled by paired, independent chains
  - Treespace visualization

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Example: Visualizing treespace with RWTY



Warren, Geneva & Lanfear (2017)

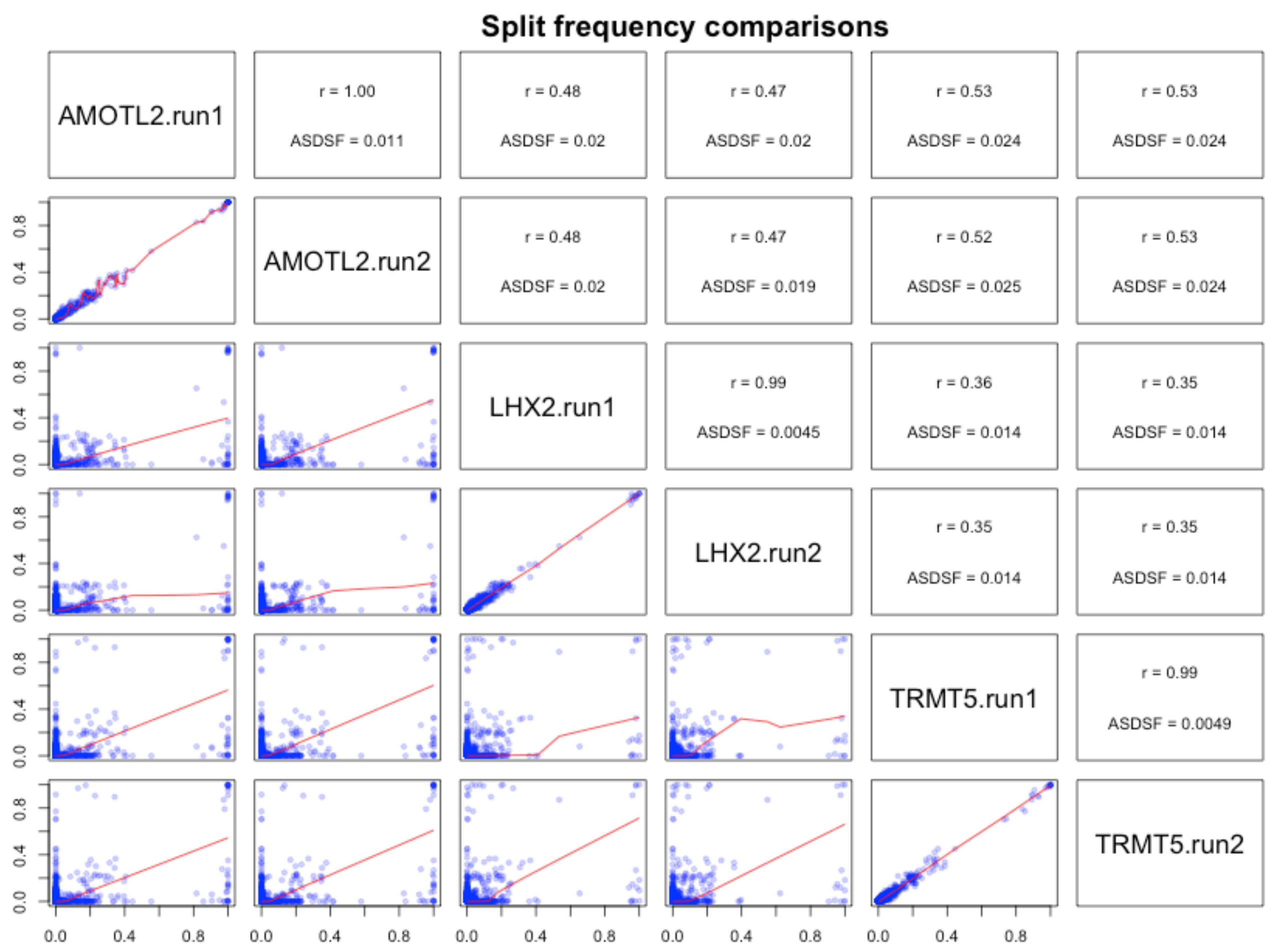
Assessing MCMC Performance: Diagnostics Based on Multiple Runs

Compare estimates from multiple independent chains

- Form of the marginal posterior densities for all parameters
  - Continuous parameters
- PSRF (Gelman–Rubin) diagnostic
  - Continuous and discrete parameters
- Comparing independent samples of trees
  - ASDSF: similarity of trees sampled by paired, independent chains
  - Treespace visualization
  - Split frequencies among runs

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

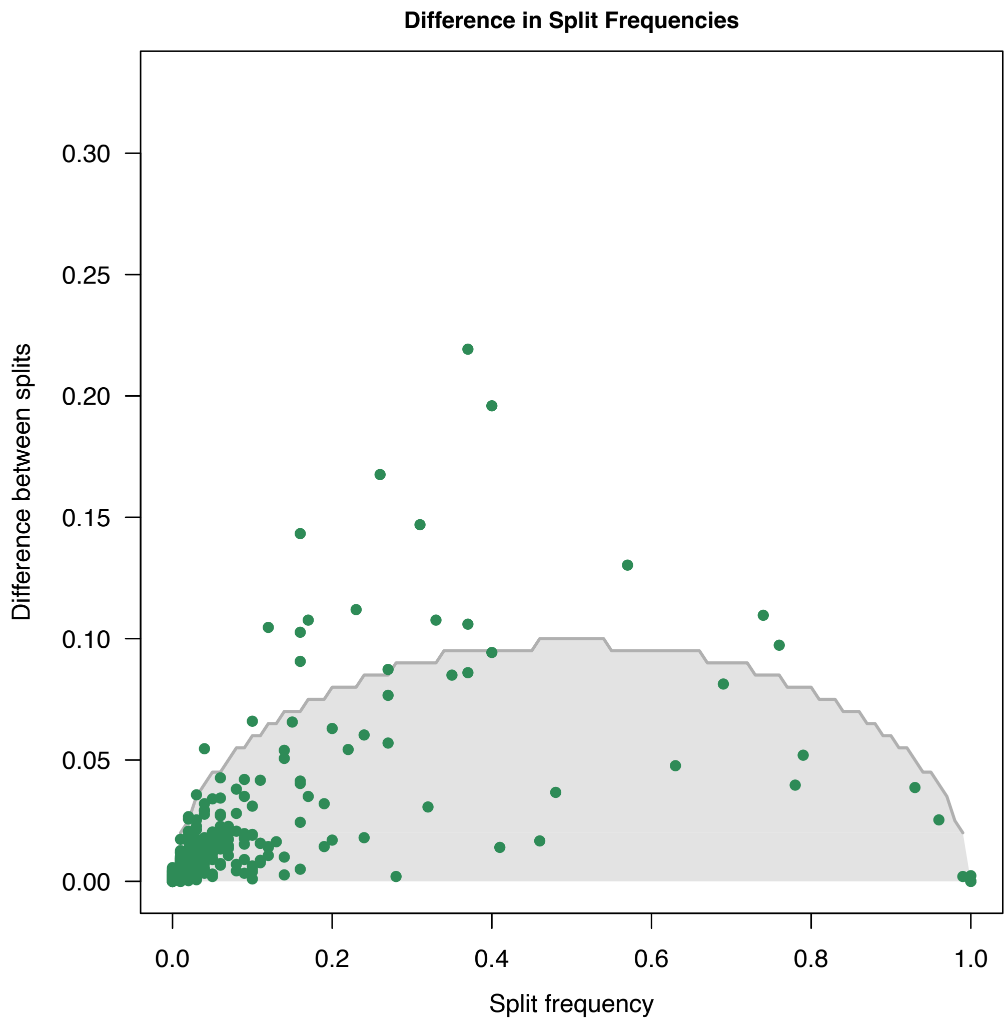
Example: Comparing split frequencies among chains



Warren, Geneva & Lanfear (2017)

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

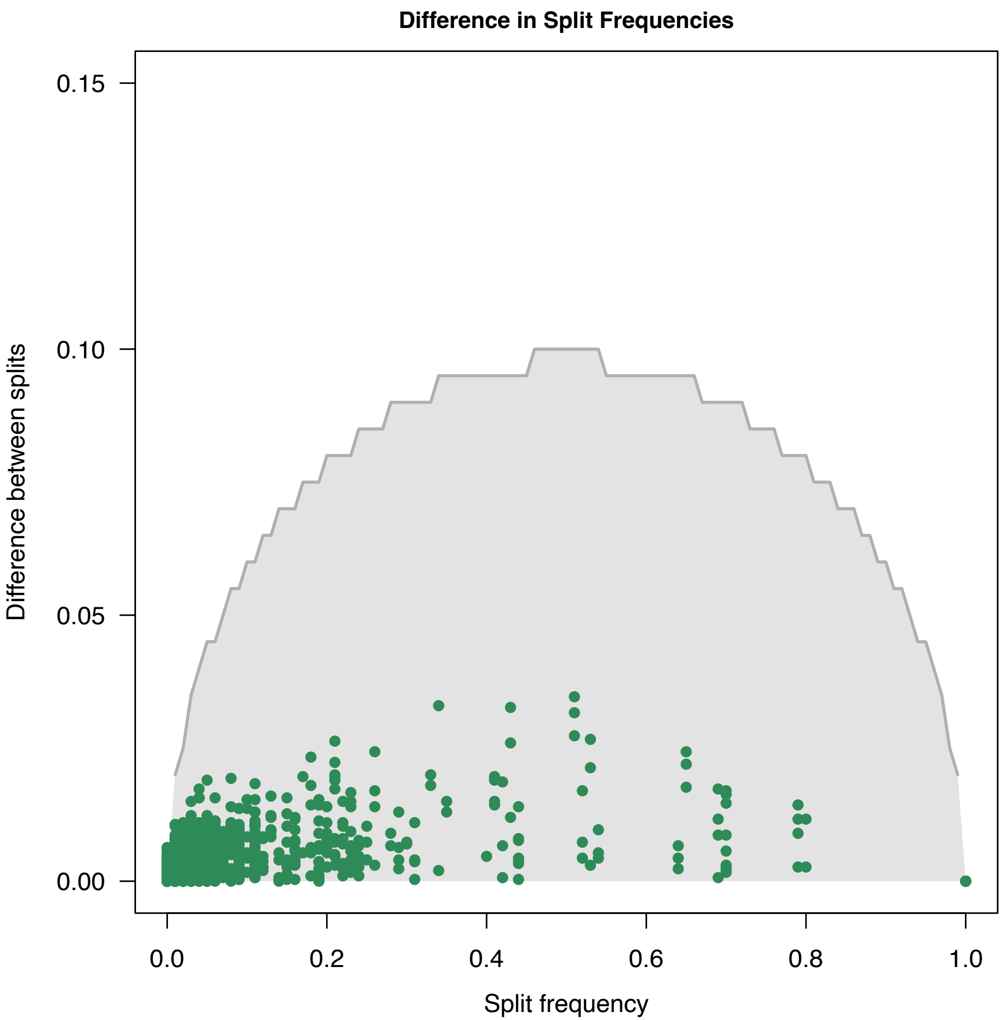
Example: Testing for equality of split frequencies among chains



Fabreti & Höhna (2022)

Assessing MCMC Performance: Diagnostics Based on Multiple Runs

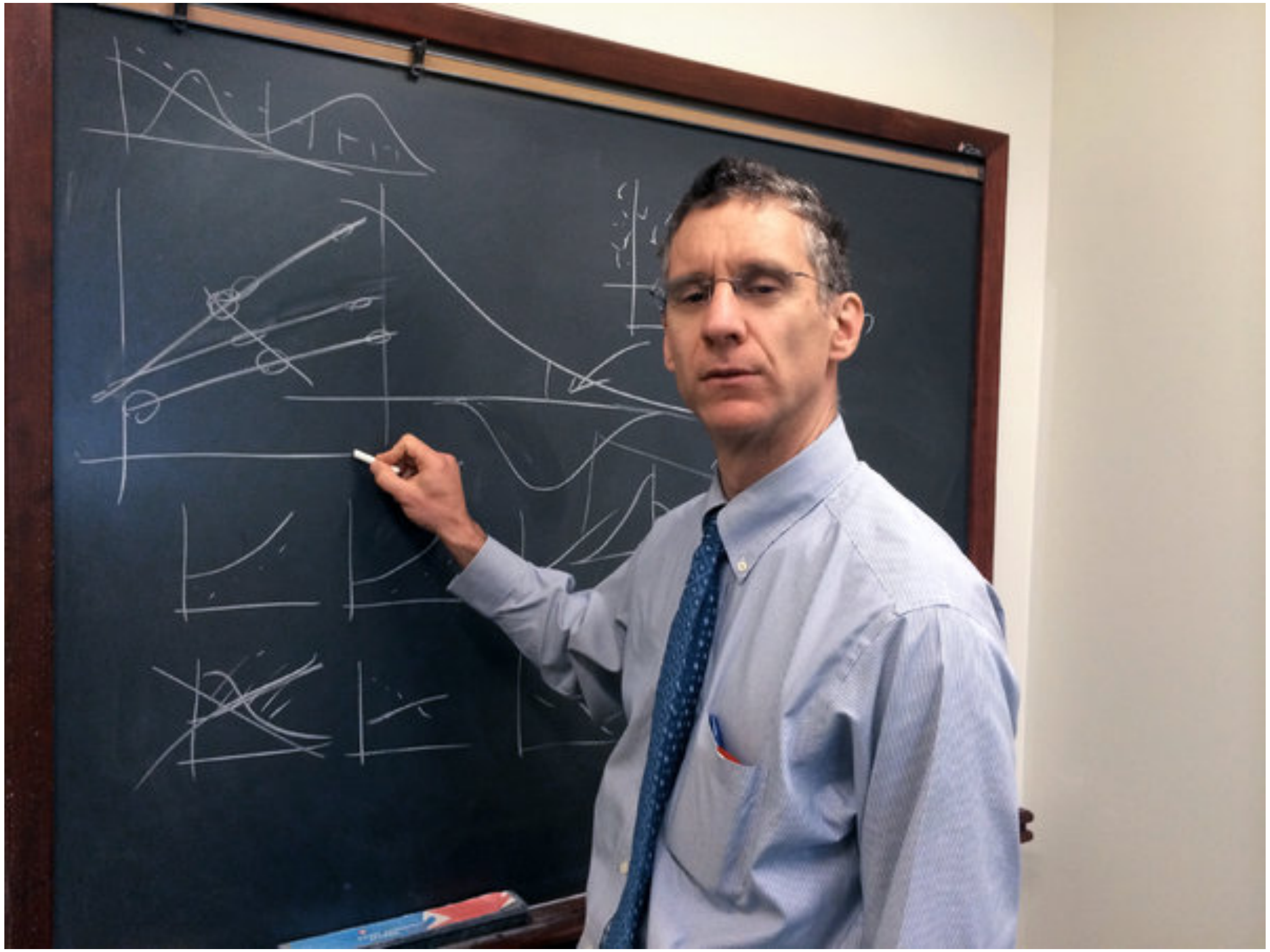
Example: Testing for equality of split frequencies among chains



Fabreti & Höhna (2022)

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." Andrew Gelman (hero)



Summary: Some General Strategies for Assessing MCMC Performance

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

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 multiple 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)

Summary: Some General Strategies for Assessing MCMC Performance

Tracer

https://github.com/beast-dev/tracer/releases/latest

RWTY

https://github.com/danlwarren/RWTY

convenience

https://revbayes.github.io/tutorials/convergence/