Bayesian Phylogenetic Inference using RevBayes:

Introduction to Bayesian Inference

Sebastian Höhna

GeoBio-Center
Ludwig-Maximilians Universität, München

Scientific Method

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

Statistical phylogenetic paradigm

what is the phylogeny of my study group?

develop a phylogenetic model with a tree (and branch lengths) and a Markov model describing how traits change over the tree

assemble a data matrix (*e.g.*, of DNA sequences) sampled from members of your study group

find the best estimate of phylogeny (and other model parameters) using a **likelihood-based** method (maximum-likelihood or Bayesian inference)

Scientific Method

Generic statistical paradigm

Statistical phylogenetic paradigm

pose a substantive question

what is the phylogeny of my study group?

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

develop a phylogenetic model with a tree (and branch lengths) and a Markov model describing how traits change over the tree

collect observations that are informative about model parameters

assemble a data matrix (*e.g.*, of DNA sequences) sampled from members of your study group

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

find the best estimate of phylogeny (and other model parameters) using a **likelihood-based** method (maximum-likelihood or Bayesian inference)

Example: Sex-ratio at birth

Table 1. SEX-RATIO OF GREY SEAL CALVES ON THE FARNE ISLANDS ACCORDING TO THEIR DATE OF BIRTH

Year	Oct. 14–27		Oct. 28–Nov. 10		Nov. 11–24		After Nov.24	
	♂	♀	♂	♀	♂	♀	♂	♀
1952	8	4	16	18	12	8	8	8
1953	12	8	11	16	5	10	3	8
1955	5	8	43	42	20	25	9	14
1956	7	6	43	47	47	47	24	32
1957	36	22	114	96	67	60	13	18
1958	1	1	57	38	45	38	11	10
1959	2	2	69	61	34	41	9	14
Total	71	51	353	318	230	229	77	104
Sex-ratio	100 : 71·8		100 : 90·1		100 : 99·6		100 : 135·1	

What is the probability of a female pup in grey seals?

Scientific Method

Generic statistical paradigm

Statistical phylogenetic paradigm

pose a substantive question

what is the phylogeny of my study group?

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

develop a phylogenetic model with a tree (and branch lengths) and a Markov model describing how traits change over the tree

collect observations that are informative about model parameters

assemble a data matrix (*e.g.*, of DNA sequences) sampled from members of your study group

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

find the best estimate of phylogeny (and other model parameters) using a **likelihood-based** method (maximum-likelihood or Bayesian inference)

Example: Sex-ratio at birth

Table 1. SEX-RATIO OF GREY SEAL CALVES ON THE FARNE ISLANDS ACCORDING TO THEIR DATE OF BIRTH

Year	Oct. 14–27		Oct. 28–Nov. 10		Nov. 11–24		After Nov.24	
	♂	♀	♂	♀	♂	♀	♂	♀
1952	8	4	16	18	12	8	8	8
1953	12	8	11	16	5	10	3	8
1955	5	8	43	42	20	25	9	14
1956	7	6	43	47	47	47	24	32
1957	36	22	114	96	67	60	13	18
1958	1	1	57	38	45	38	11	10
1959	2	2	69	61	34	41	9	14
Total	71	51	353	318	230	229	77	104
Sex-ratio	100 : 71·8		100 : 90·1		100 : 99·6		100 : 135·1	

$$P(male) = p$$

$$P(female) = 1 - p$$

$$P(x = 731, n = 1433) = \binom{1433}{731} p^{731}(1-p)^{1433-713}$$

Scientific Method

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

Statistical phylogenetic paradigm

what is the phylogeny of my study group?

develop a phylogenetic model with a tree (and branch lengths) and a Markov model describing how traits change over the tree

assemble a data matrix (*e.g.*, of DNA sequences) sampled from members of your study group

find the best estimate of phylogeny (and other model parameters) using a **likelihood-based** method (maximum-likelihood or Bayesian inference)

Bayesian vs Frequentist

Bayesian:

Frequentist:

Bayesian vs Frequentist

Bayesian:
- Condition on observing the data.

Are we asking how the data and parameter estimates behave under repetition (e.g., in simulations or mass-production) or asking about one/few unique observations?

- Estimate posterior probability of parameter.

Both methods are model based inferences and use a likelihood function.

- The posterior distribution provides a credible interval.

The maximum likelihood estimate is good if you want a single estimate, the credible interval is good if you want a measure uncertainty.

- Assume that each parameter has prior probability. Parameters are treated similar to observations.

Frequentist:
- Assume data was observed from repeated experiments.

- Find parameter value with highest likelihood.
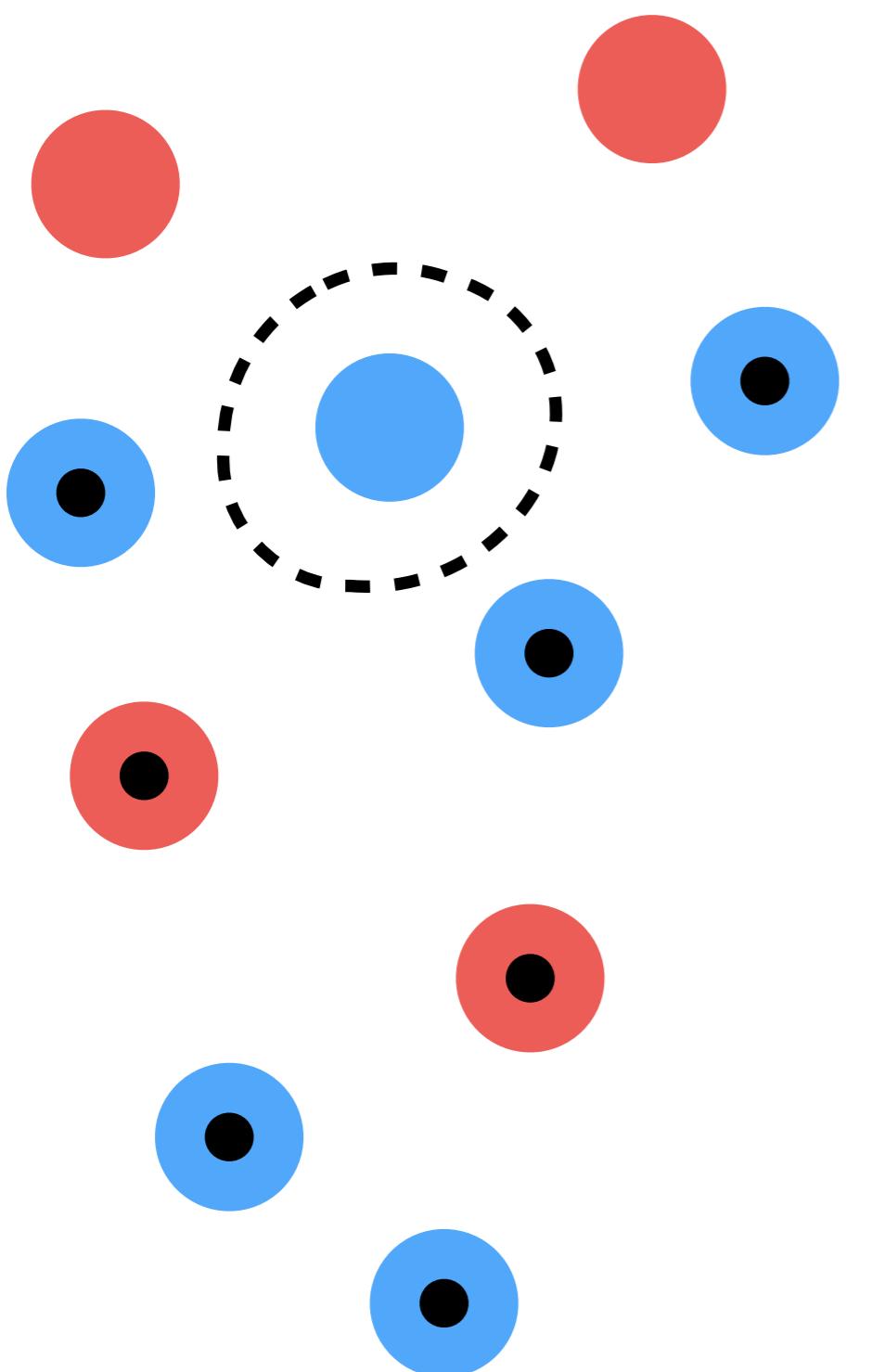
- Likelihood can be used to generate a confidence interval.

- There is only one true parameter. Parameters do not have a probability distribution and are not random variables!
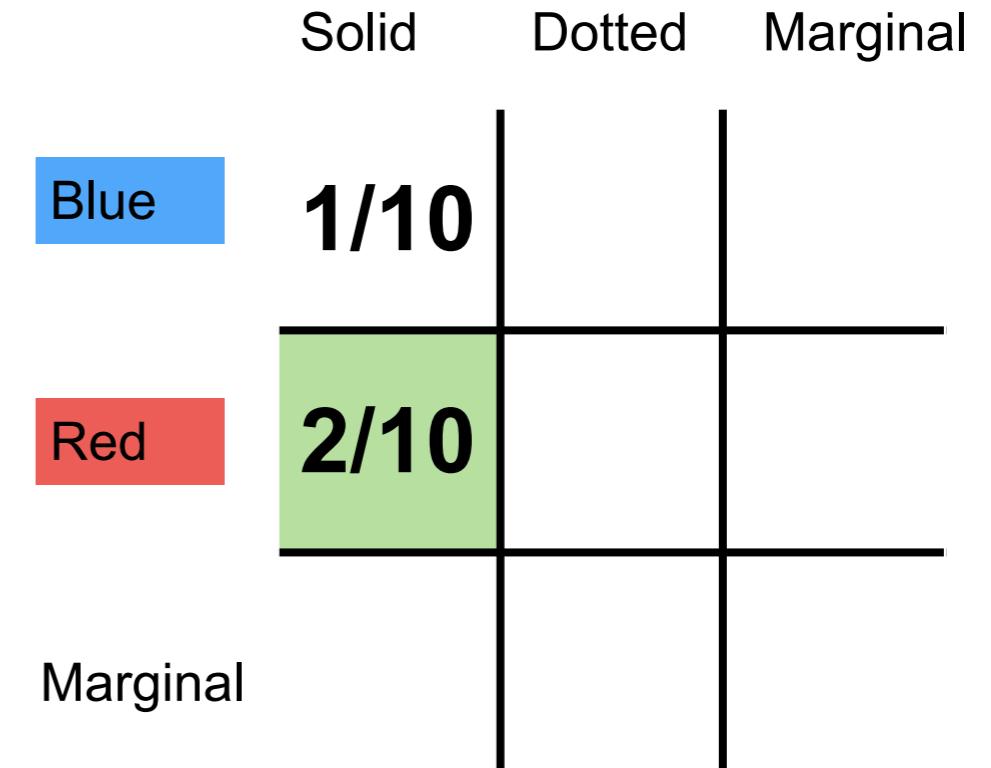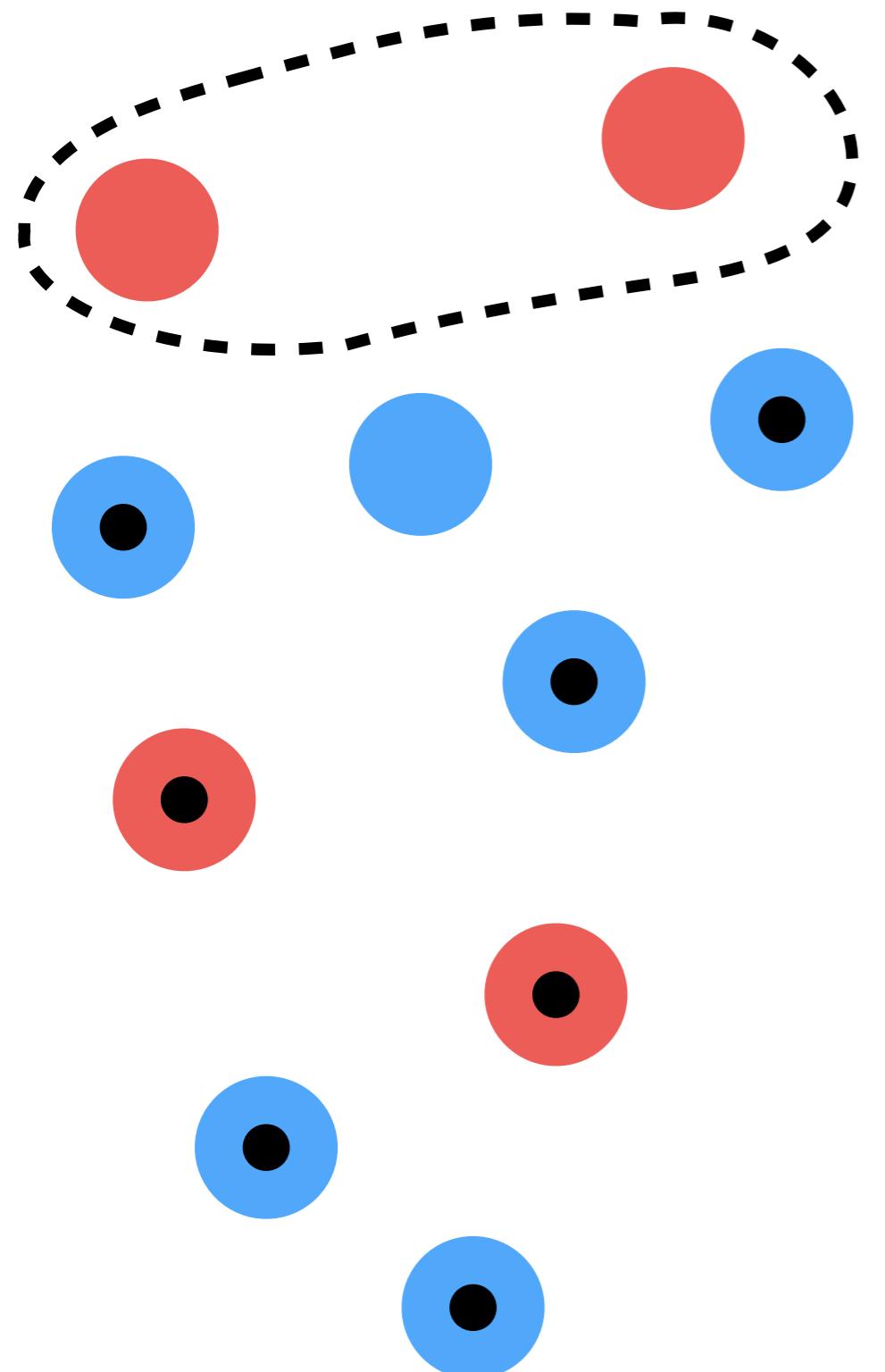
Choosing priors can be difficult but is very important.
Priors allow us to build hierarchical models easily.

Excursus: Primer on Probabilities

Excursus: Primer on Probabilities

Excursus: Primer on Probabilities

	Solid	Dotted	Marginal
Blue	1/10		
Red	2/10		
Marginal			

$$P(\text{●}) = P(B,S) = 1/10$$

$$P(\text{●}) = P(R,S) = 2/10$$
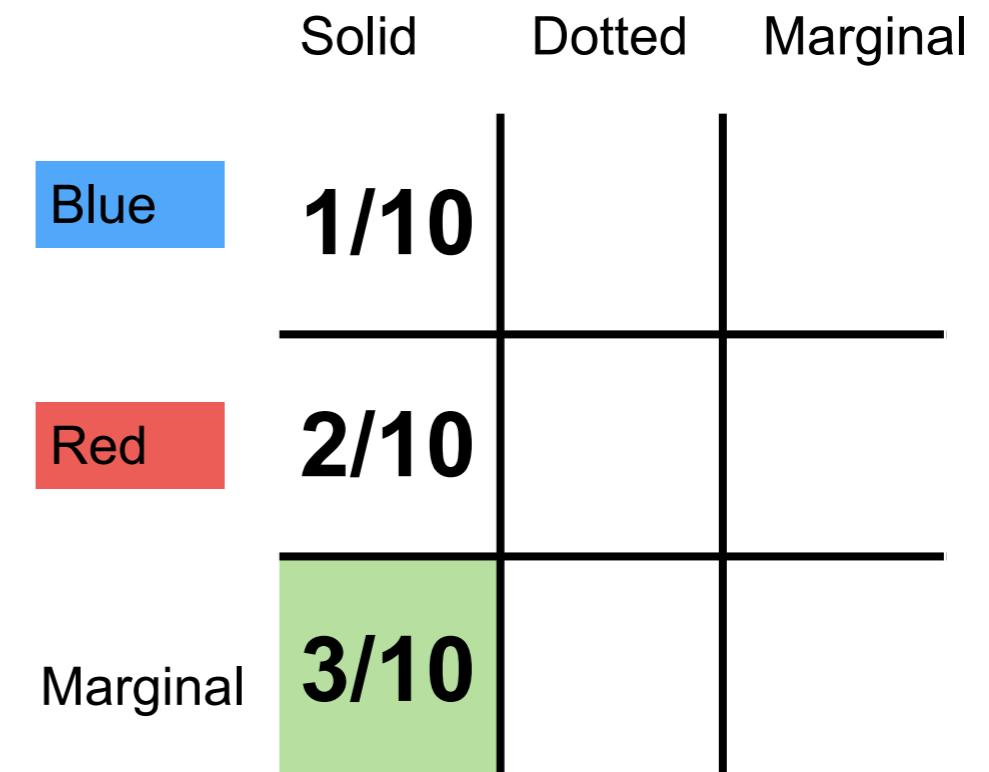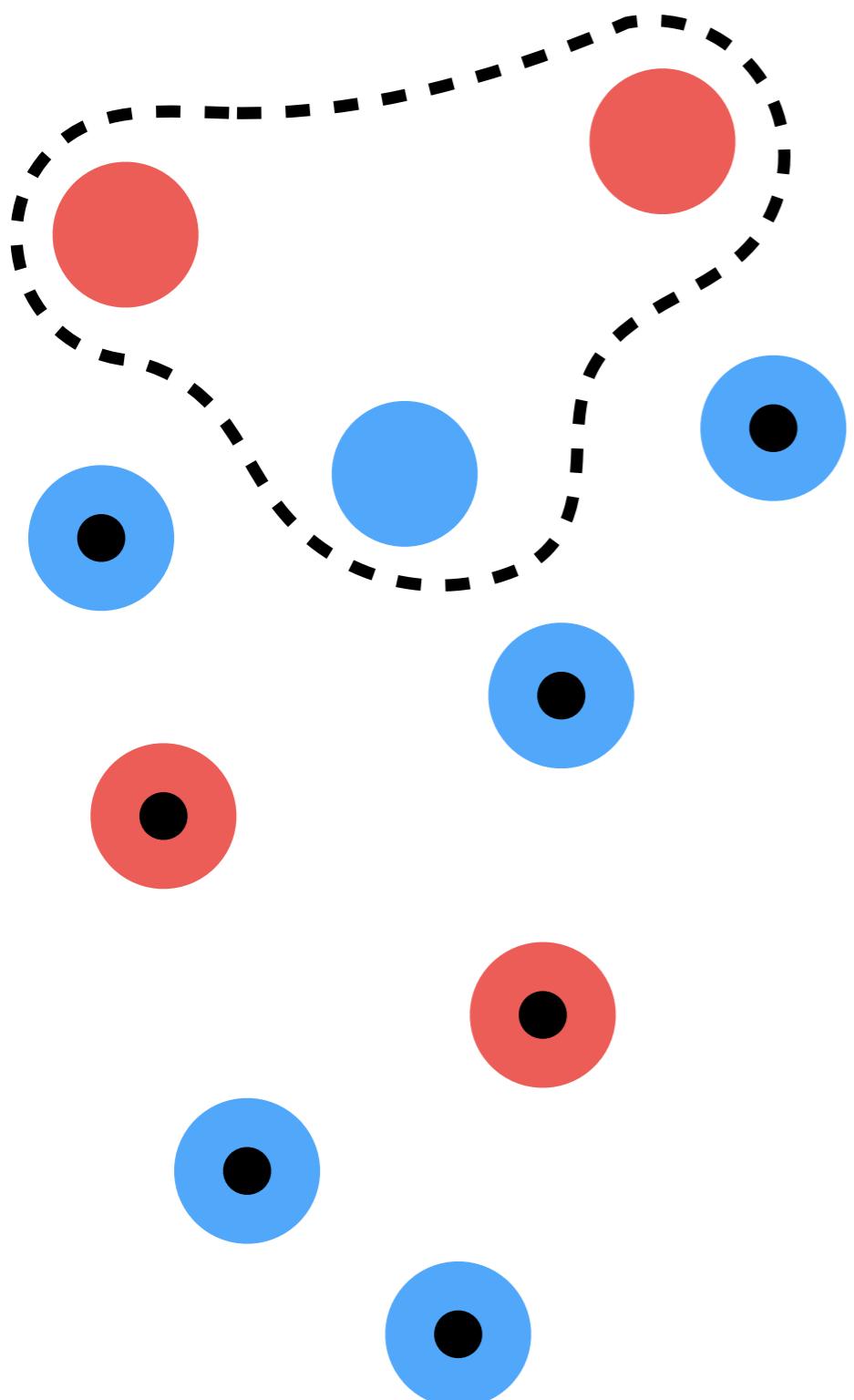
Excursus: Primer on Probabilities

	Solid	Dotted	Marginal
Blue	**1/10**		
Red	**2/10**		
Marginal	**3/10**		
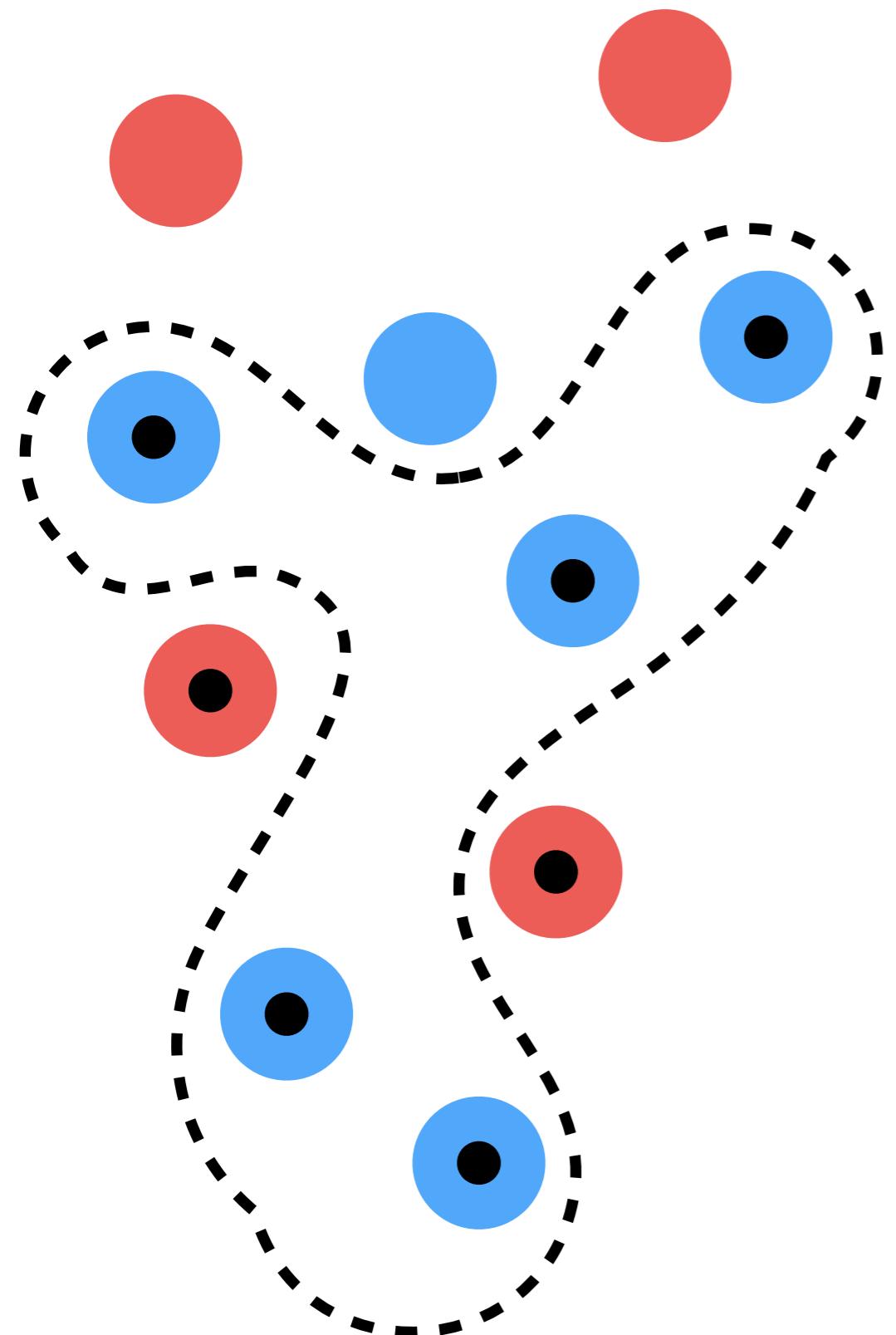
$$P(\ \bullet\ ) = P(B,S) = 1/10$$

$$P(\ \bullet\ ) = P(R,S) = 2/10$$

Excursus: Primer on Probabilities

Excursus: Primer on Probabilities
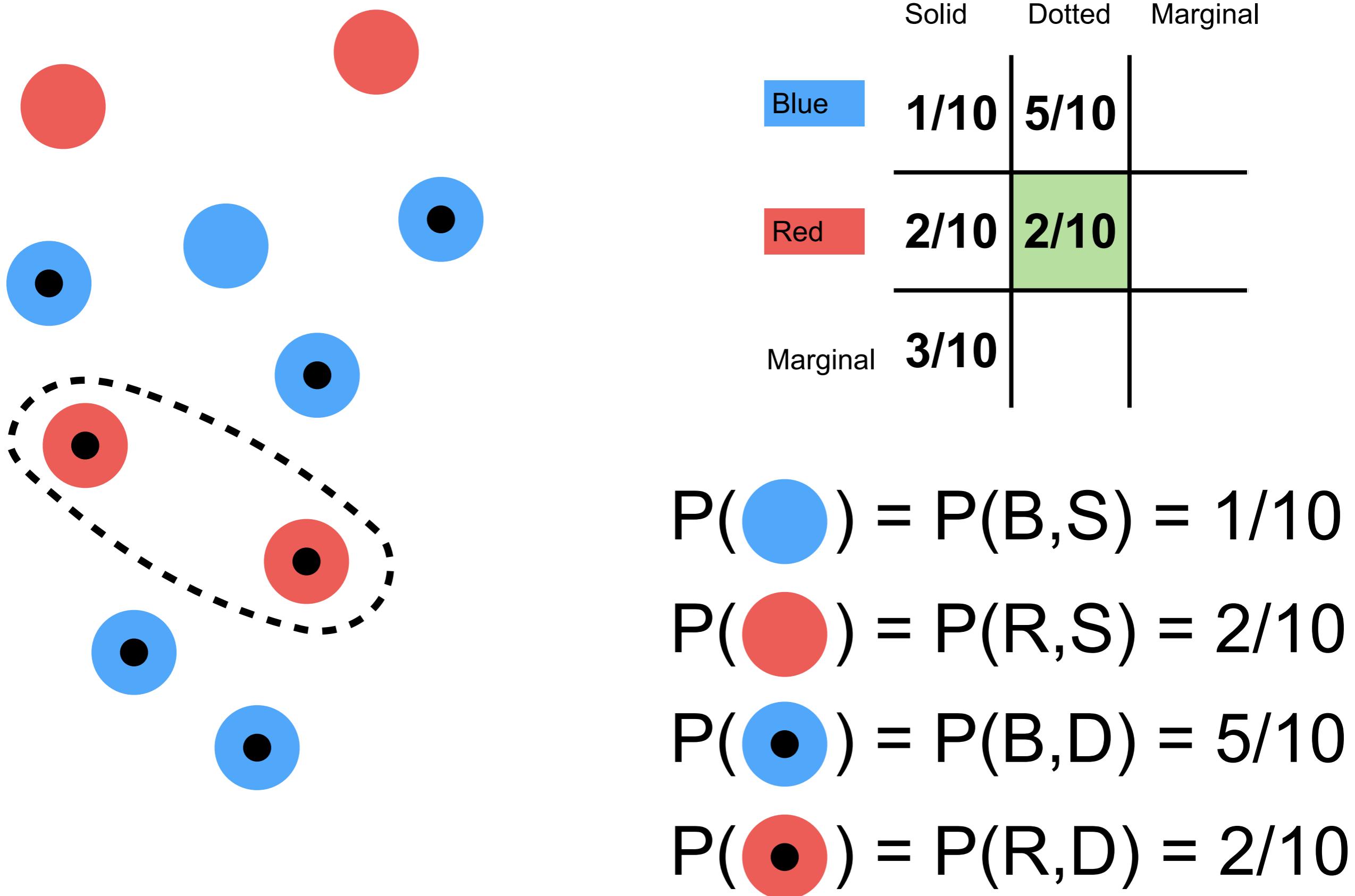
|	Solid	Dotted	Marginal
Blue	**1/10**	**5/10**	
Red	**2/10**	**2/10**	
Marginal	**3/10**		

$$P(\text{🔵}) = P(B,S) = 1/10$$

$$P(\text{🔴}) = P(R,S) = 2/10$$

$$P(\text{🔵•}) = P(B,D) = 5/10$$

$$P(\text{🔴•}) = P(R,D) = 2/10$$

Excursus: Primer on Probabilities

Excursus: Primer on Probabilities

	Solid	Dotted	Marginal
Blue	**1/10**	**5/10**	**6/10**
Red	**2/10**	**2/10**	
Marginal	**3/10**	**7/10**	

$P(\ \bullet\ ) = P(B,S) = 1/10$

$P(\ \bullet\ ) = P(R,S) = 2/10$

$P(\ \bullet\ ) = P(B,D) = 5/10$

$P(\ \bullet\ ) = P(R,D) = 2/10$

Excursus: Primer on Probabilities

Excursus: Primer on Probabilities

Joint Probabilities

P(🔵) = P(B,S) = 1/10

P(🔴) = P(R,S) = 2/10

P(🔵•) = P(B,D) = 5/10

P(🔴•) = P(R,D) = 2/10

	Solid	Dotted	Marginal
Blue	1/10	5/10	6/10
Red	2/10	2/10	4/10
Marginal	3/10	7/10	

Excursus: Primer on Probabilities

Joint Probabilities

P(🔵) = P(B,S) = 1/10

P(🔴) = P(R,S) = 2/10

P(🔵•) = P(B,D) = 5/10

P(🔴•) = P(R,D) = 2/10

	Solid	Dotted	Marginal
Blue	**1/10**	**5/10**	**6/10**
Red	**2/10**	**2/10**	**4/10**
Marginal	**3/10**	**7/10**	

Marginal Probabilities

P(🔵) + P(🔵•) = P(B) = 6/10

P(🔴) + P(🔴•) = P(R) = 4/10

P(🔵) + P(🔴) = P(S) = 3/10

P(🔵•) + P(🔴•) = P(D) = 7/10

Excursus: Primer on Probabilities

Joint Probabilities

P(🔵) = P(B,S) = 1/10

P(🔴) = P(R,S) = 2/10

P(🔵•) = P(B,D) = 5/10

P(🔴•) = P(R,D) = 2/10

	Solid	Dotted	Marginal
Blue	**1/10**	**5/10**	**6/10**
Red	**2/10**	**2/10**	**4/10**
Marginal	**3/10**	**7/10**	

Marginal Probabilities

P(🔵) + P(🔵•) = P(B) = 6/10

P(🔴) + P(🔴•) = P(R) = 4/10

P(🔵) + P(🔴) = P(S) = 3/10

P(🔵•) + P(🔴•) = P(D) = 7/10

Conditional Probabilities

P(S|R) = P(R,S) / P(R)

$$= \frac{P(R,S)}{P(R,S) + P(R,D)}$$

$$= \frac{P(🔴)}{P(🔴) + P(🔴•)}$$

Bayes' Rule

P(Parameters | Data) = P(Parameters | Data)
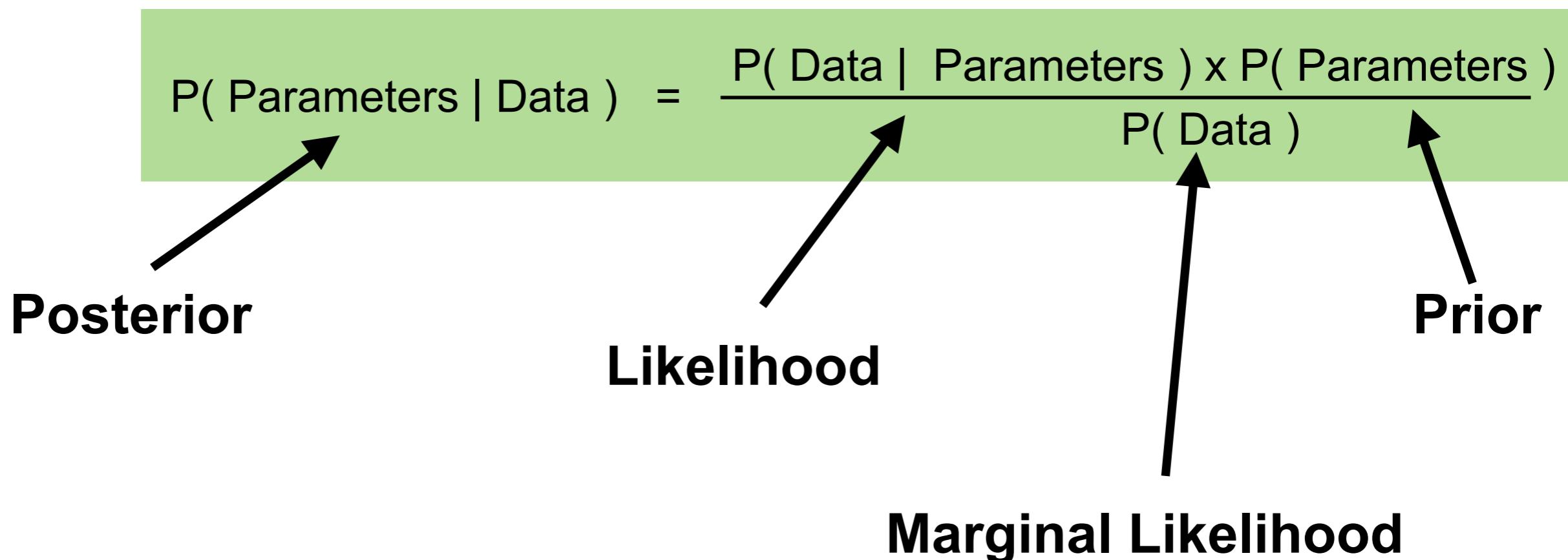
Multiply with P(Data):

$$= \frac{P(\text{ Parameters } | \text{ Data }) \times P(\text{ Data })}{P(\text{ Data })}$$

Apply rule of conditional probability:

$$= \frac{P(\text{ Parameters, Data })}{P(\text{ Data })}$$

Again, apply rule of conditional probability:

$$P(\text{ Parameters } | \text{ Data }) = \frac{P(\text{ Data } | \text{ Parameters }) \times P(\text{ Parameters })}{P(\text{ Data })}$$

Posterior

Likelihood

Prior

Marginal Likelihood
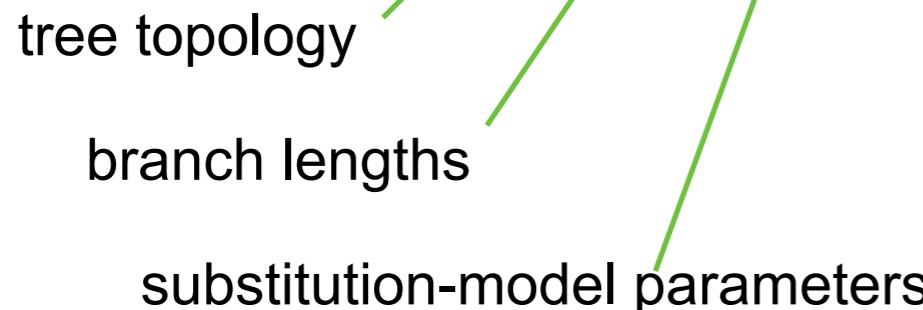
Likelihood Function

The likelihood function is defined as the probability of observing the data, **X**, as a function of the model parameters, Θ.

$$L(\Theta) = f(X, \Theta)$$

For example, in phylogenetic tree estimation the likelihood is defined as the probability of observing the sequence alignment, **X**, under a fully specified phylogenetic model.

$$L(\tau, \nu, \Phi) = P(X|\tau, \nu, \Phi)$$

tree topology

branch lengths

substitution-model parameters

Scientific Method

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

Statistical phylogenetic paradigm

what is the phylogeny of my study group?

develop a phylogenetic model with a tree (and branch lengths) and a Markov model describing how traits change over the tree

assemble a data matrix (*e.g.*, of DNA sequences) sampled from members of your study group

find the best estimate of phylogeny (and other model parameters) using a **likelihood-based** method (maximum-likelihood or Bayesian inference)

Example: Sex-ratio at birth (continued)

Prior:

$$P(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

Likelihood:

$$P(x|p) = \binom{m+f}{m} p^m (1-p)^f$$

Posterior:

$$P(p|x) = \frac{P(x|p) \times P(p)}{P(x)}$$
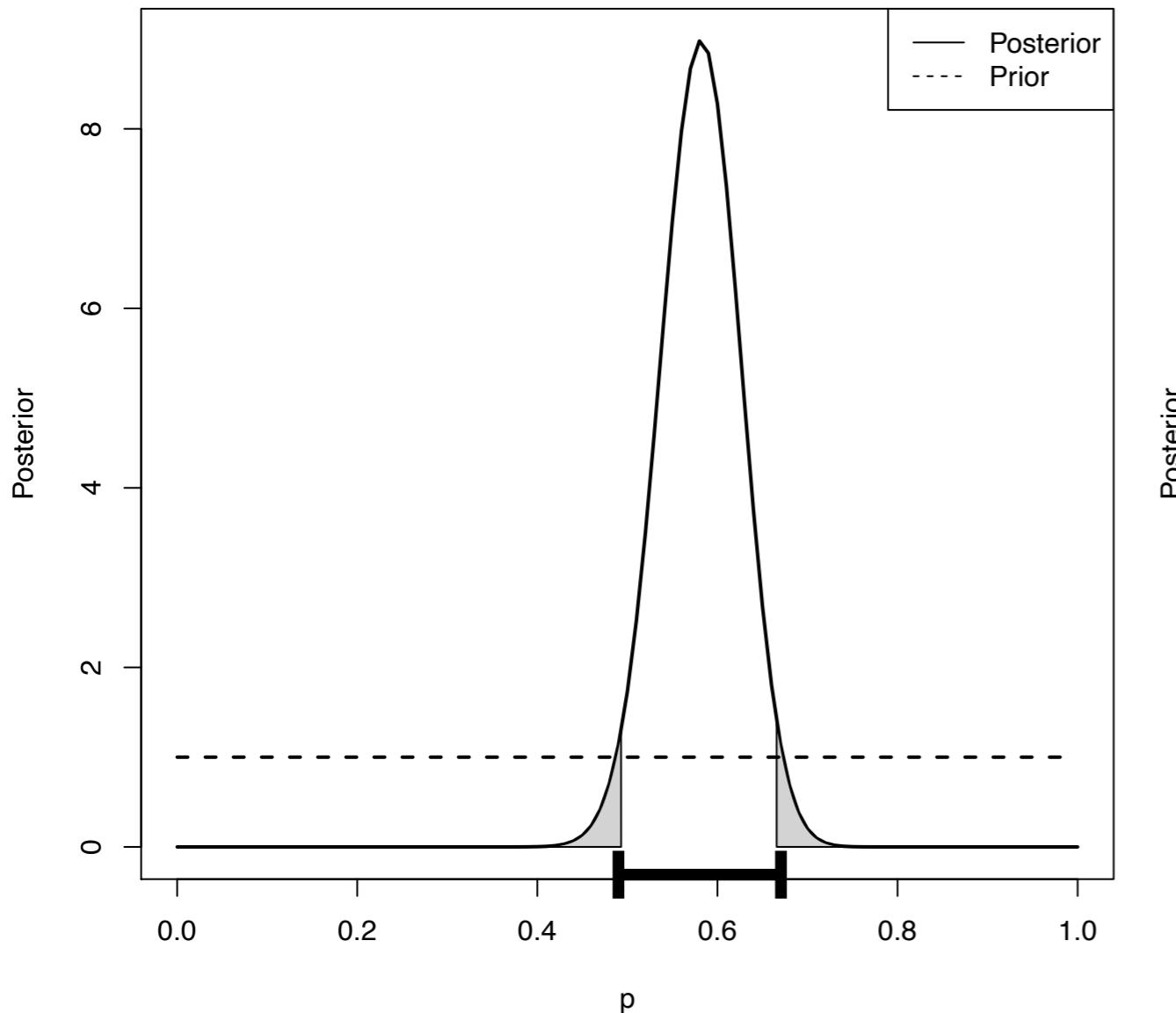
$$P(p|x) = \frac{P(x|p) \times P(p)}{\int_0^1 P(x|p) \times P(p)\,dp}$$

$$P(p|x) = \frac{\binom{m+f}{m} p^{m+\alpha-1}(1-p)^{f+\beta-1}\frac{1}{B(\alpha,\beta)}}{\int_0^1 \left(\binom{m+f}{m} p^{m+\alpha-1}(1-p)^{f+\beta-1}\frac{1}{B(\alpha,\beta)}\right)dp}$$
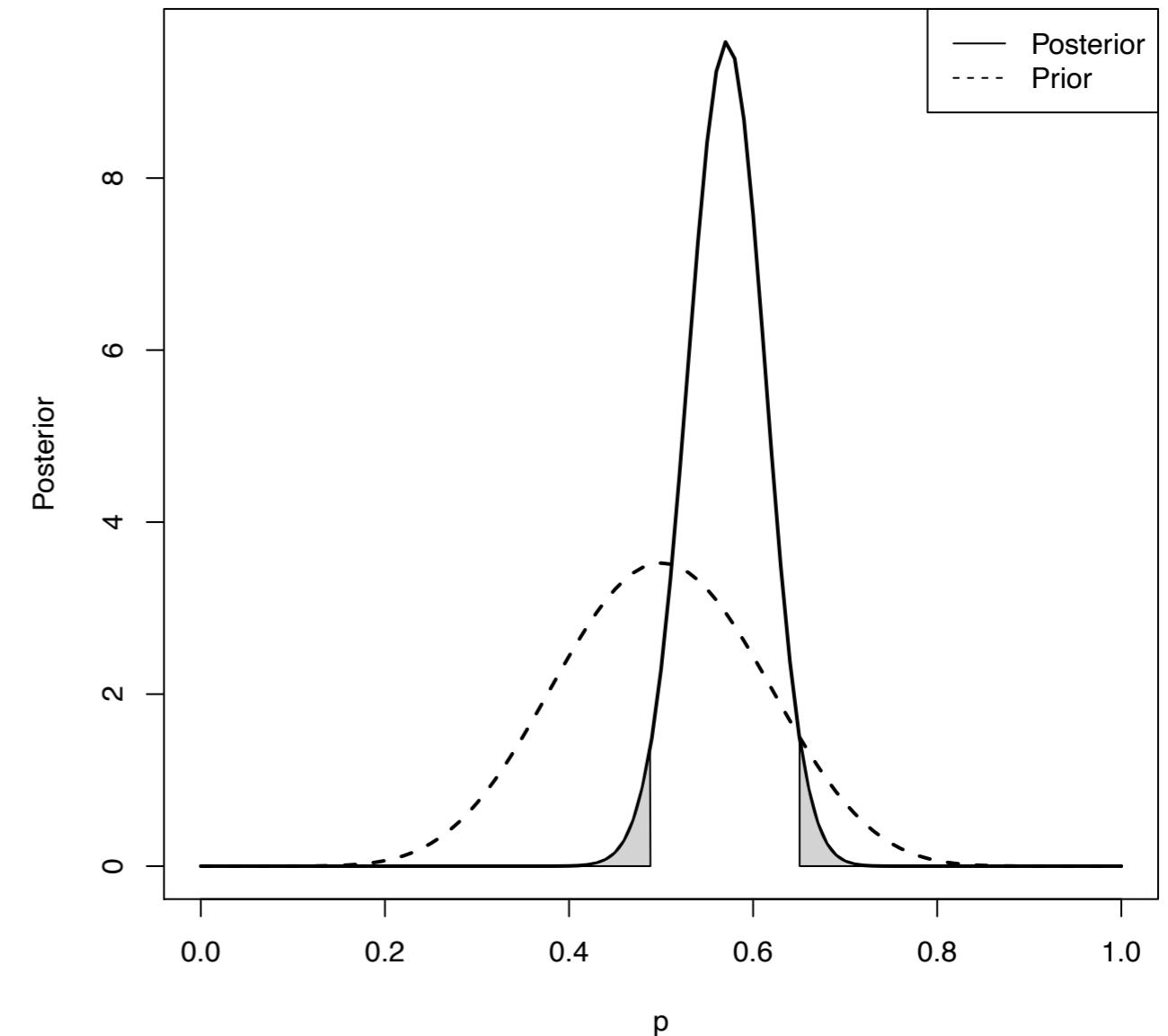
$$P(p|x) = \frac{p^{m+\alpha-1}(1-p)^{f+\beta-1}}{\int_0^1 \left(p^{m+\alpha-1}(1-p)^{f+\beta-1}\right)dp} \longrightarrow \boxed{Beta(m+\alpha, f+\beta)}$$

Example: Sex-ratio at birth (continued)

Prior: Beta(1,1)

Prior: Beta(10,10)
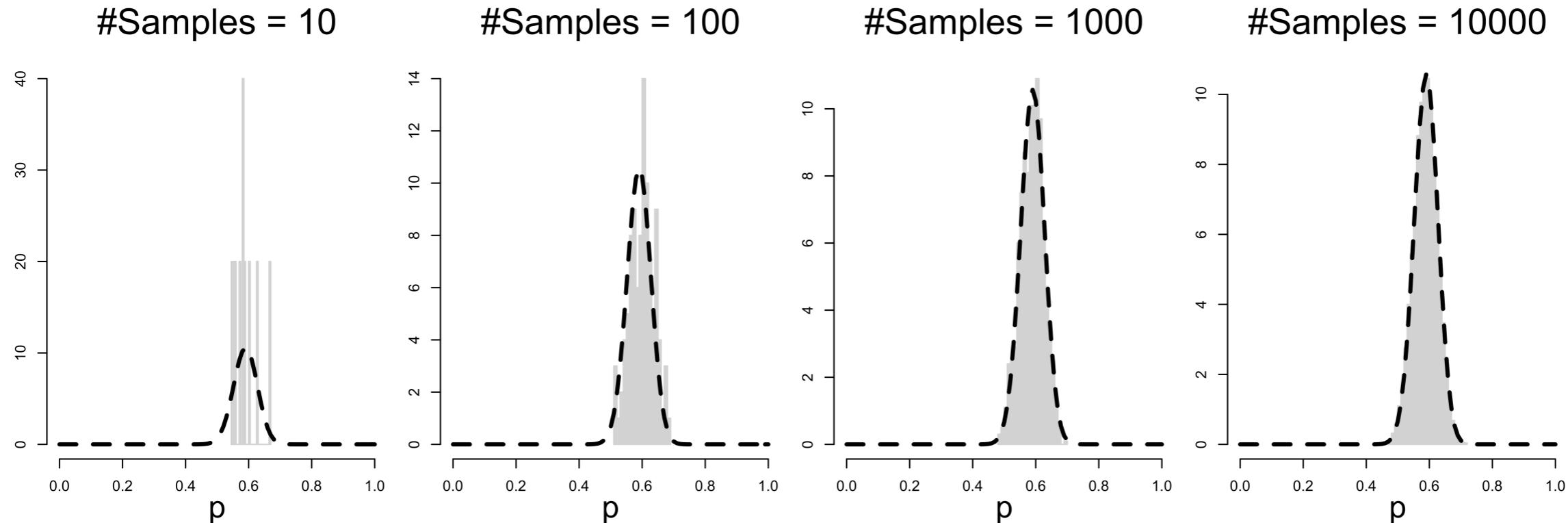
95% HPD (Highest Posterior Density)

Approximating the Posterior Probability using
Markov chain Monte Carlo sampling

What if you don't know the analytical solution for the posterior distribution?

Use the Metropolis-Hastings algorithm to sample from the posterior distribution!

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value, θ', for the selected parameter.

4. Calculate the probability R of accepting the move.
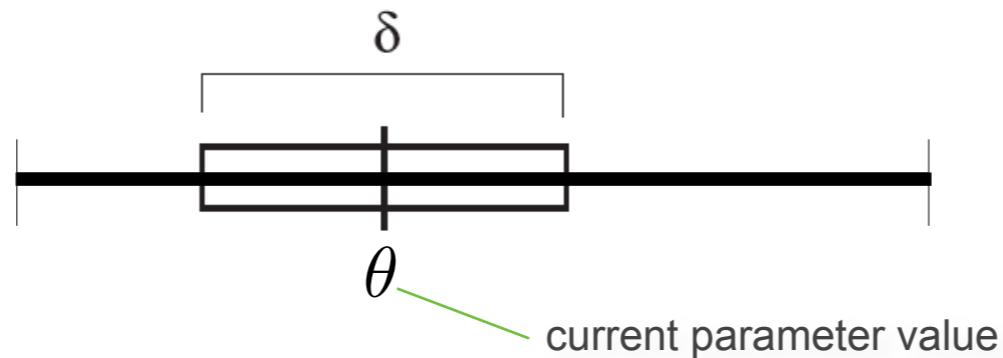
$$R = \min\left[1, \underbrace{\frac{f(\mathbf{X}|\theta')}{f(\mathbf{X}|\theta)}}_{\substack{\text{Likelihood} \\ \text{Ratio}}} \cdot \underbrace{\frac{f(\theta')}{f(\theta)}}_{\substack{\text{Prior} \\ \text{Ratio}}} \cdot \underbrace{\frac{f(\theta|\theta')}{f(\theta'|\theta)}}_{\substack{\text{Proposal} \\ \text{Ratio}}}\right]$$

5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u$.

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

Sliding Window MCMC Move

Sliding window move

1. Draw a random variable $u \sim \mathrm{Uniform}(-\delta, \delta)$
2. Update θ by the amount of u

current parameter value

- the proposal density is controlled by the tuning parameter, δ
- when δ is large, larger changes will be proposed
- when δ is small, smaller changes will be proposed

Works well for location parameters.
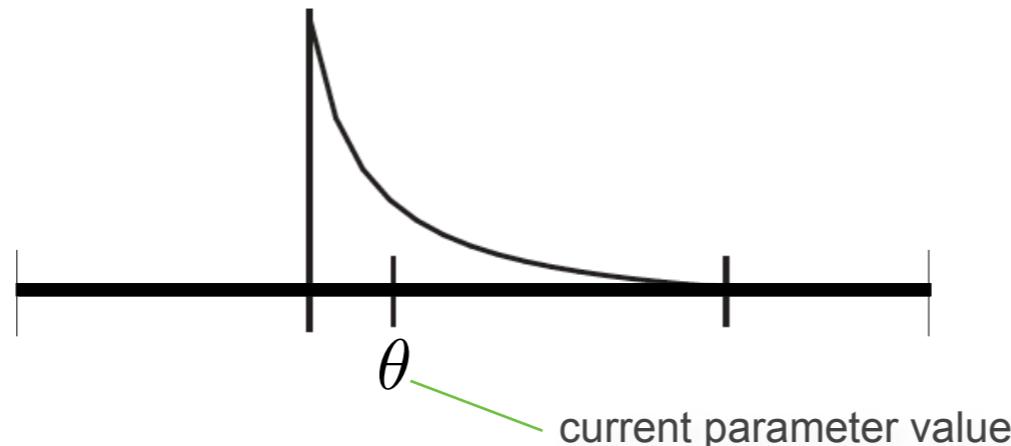
```
theta ~ dnNormal( mean, standard_deviation )
moves.append( mvSlide(epsilon, delta=0.8, tune=true, weight=3.0) )
```

Scaling (multiplier) MCMC Move

Scaling m

1. Draw a random variable $u \sim \text{Uniform}(-\lambda, \lambda)$
2. Multiply θ by the amount of e^u
3. Compute the Hastings ratio as $HR = e^u$

θ — current parameter value

- the proposal density is controlled by the tuning parameter, δ
- when δ is large, larger changes will be proposed
- when δ is small, smaller changes will be proposed

Equivalent to sliding-window proposal with log-transformed x axis

Works well for rate parameters.
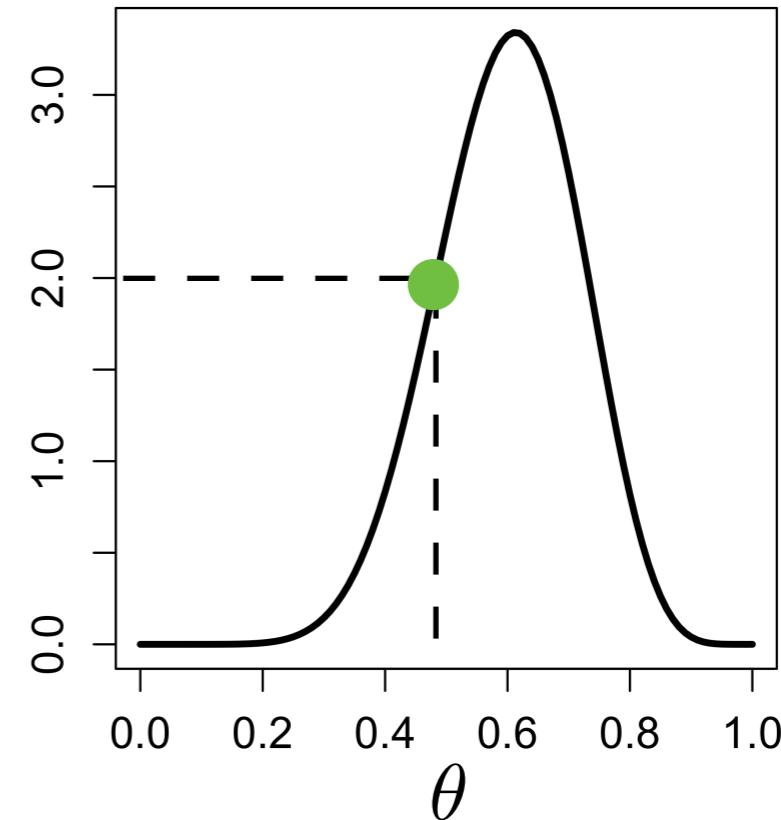
```
br_lens[i] ~ dnExponential(10.0)
moves.append( mvScale(br_lens[i],lambda=1,tune=true,weight=1) )
```

Approximating the Posterior Probability using Markov chain Monte Carlo sampling

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value,   , for the selected parameter.

4. Calculate the probability R of accepting the move.

5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u.$

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value, , for the selected parameter.

4. Calculate the probability R of accepting the move.

5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u$.

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value, , for the selected parameter.

4. Calculate the probability R of accepting the move.

5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u$.

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

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

likelihood ratio   prior ratio   proposal ratio

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value,   , for the selected parameter.

4. Calculate the probability R of accepting the move.

5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u$.

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

$$R = \min \left[1, \frac{f(\theta'|x)}{f(\theta|x)} \right] = \frac{3}{2}$$

Approximating the Posterior Probability using Markov chain Monte Carlo sampling

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.
2. Select a parameter to update according to some proposal mechanism (i.e., move).
3. Propose a new value, , for the selected parameter.
4. Calculate the probability R of accepting the move.
5. Generate a uniform random variable, u ~ Uniform(0,1), and accept if $R > u$.
6. Repeat step 2-5 and store the parameter values to a file every k iterations.

$$R = \min \left[1, \frac{f(\theta'|x)}{f(\theta|x)} \right] = \frac{3}{2}$$

ACCEPT!

Approximating the Posterior Probability using Markov chain Monte Carlo sampling

The Metropolis-Hastings algorithm

1. Initialize the chain with some random values for all parameters, e.g., the tree with branch length.

2. Select a parameter to update according to some proposal mechanism (i.e., move).

3. Propose a new value, , for the selected parameter.

4. Calculate the probability R of accepting the move.

5. Generate a uniform random variable, $u \sim$ Uniform(0,1), and accept if $R > u.$

6. Repeat step 2-5 and store the parameter values to a file every k iterations.

$$R = \min \left[1, \frac{f(\theta'|x)}{f(\theta|x)} \right] = \frac{1}{3}$$

REJECT!

MCMC Robot

MCMC demonstration software developed by Paul Lewis.

Free Windows and iPhone App: **http://www.mcmcrobot.org or https://phylogeny.uconn.edu/software/#**
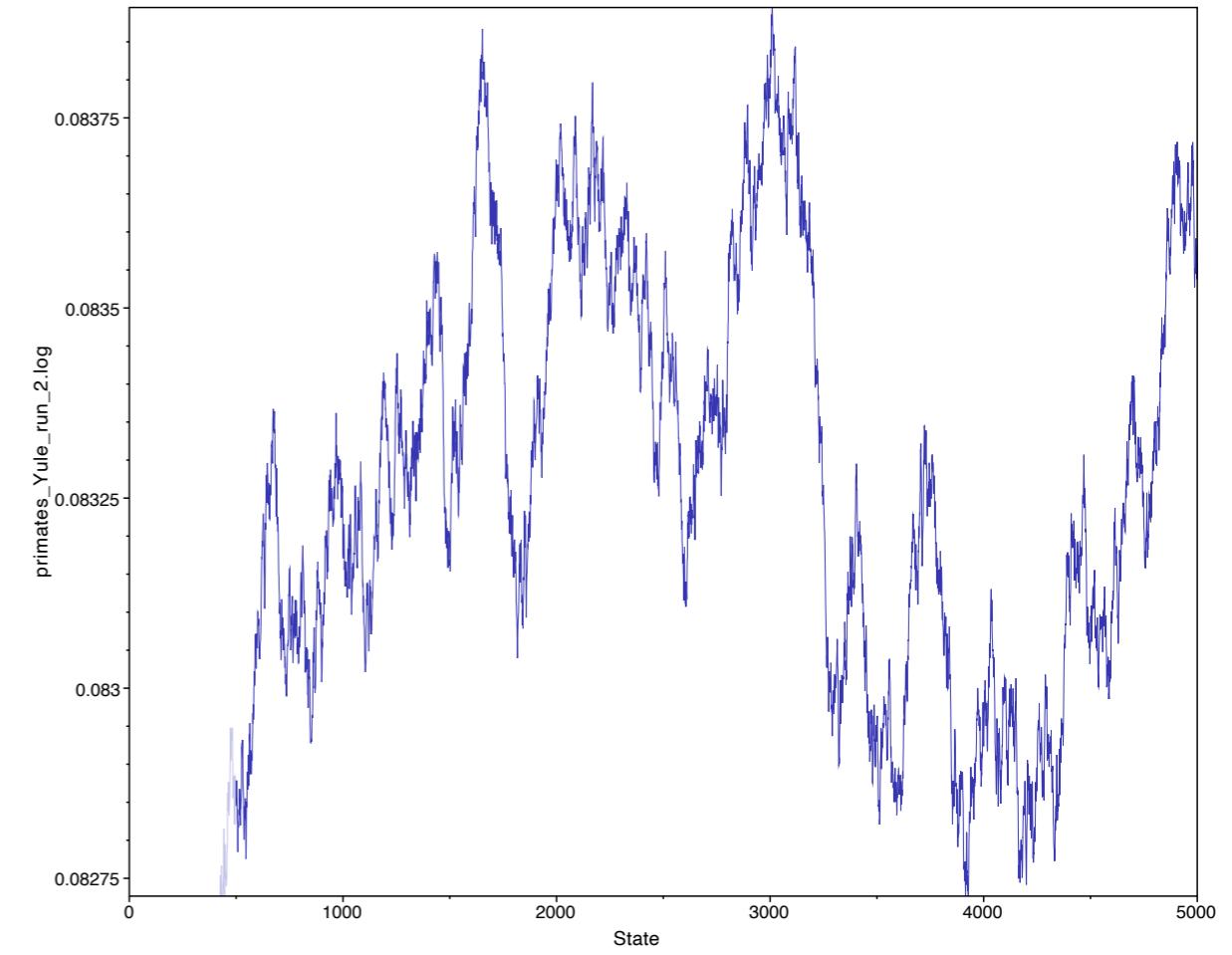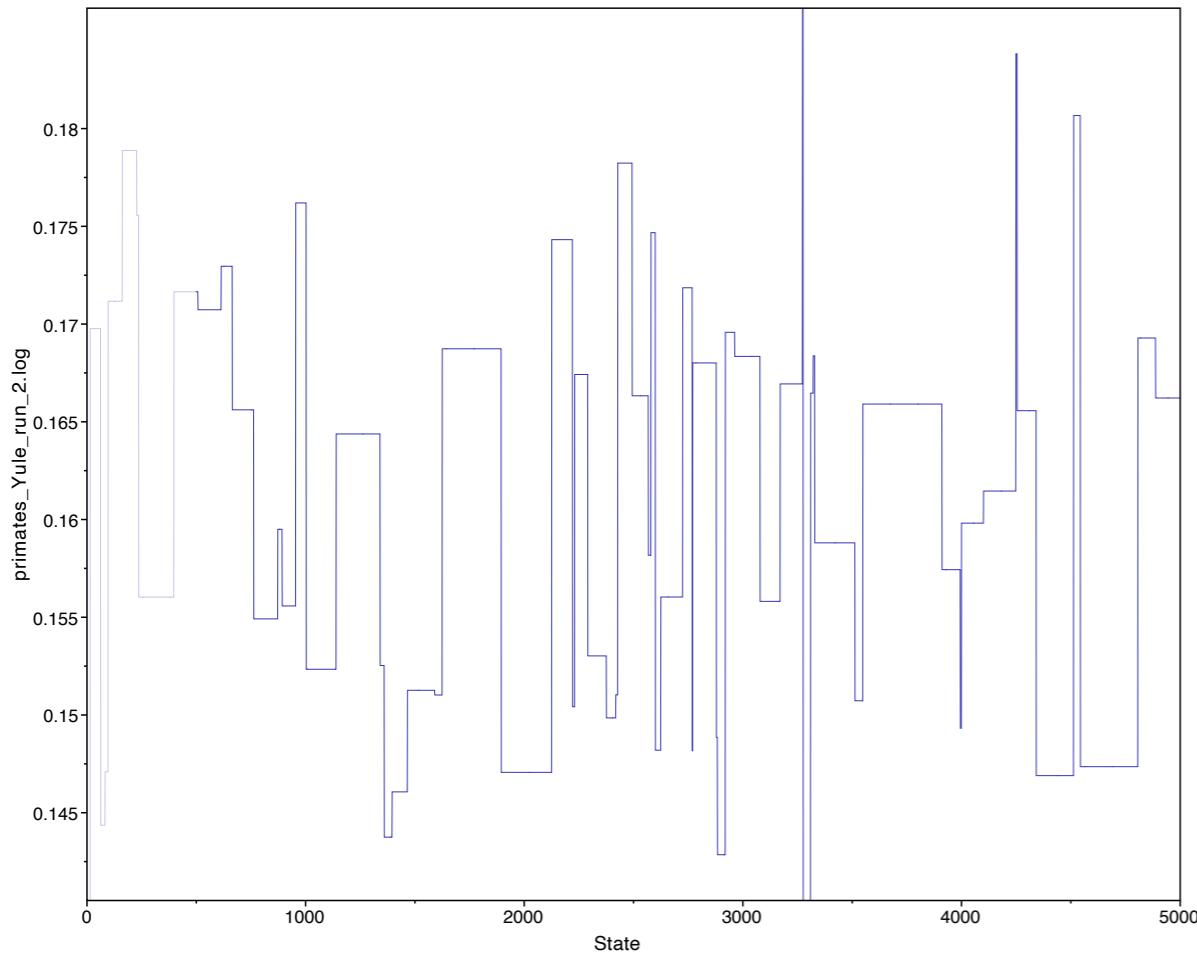
Same idea for Mac developed by John Huelsenbeck.
**http://cteg.berkeley.edu/software.html**

Effective Sample Size
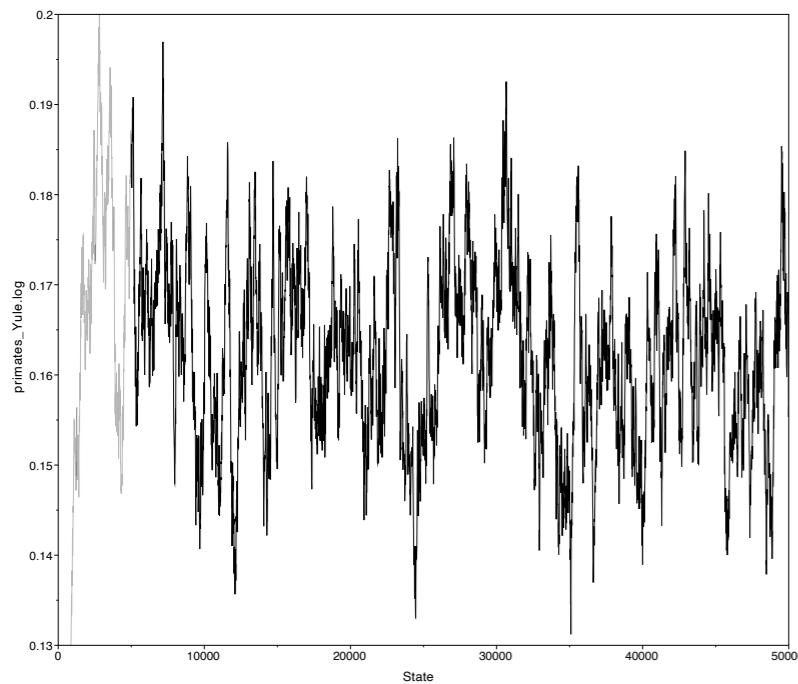
These are MCMC runs with 5000 iterations. **How many samples from the posterior do these represent?**
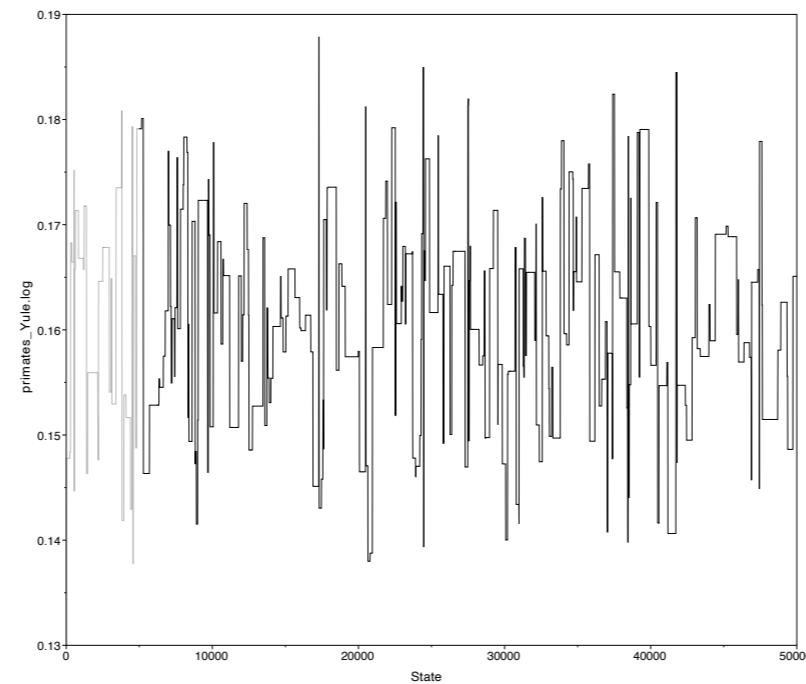

The number of independent samples equivalent to these correlated samples is called the **effective sample size.**

Assessing Convergence
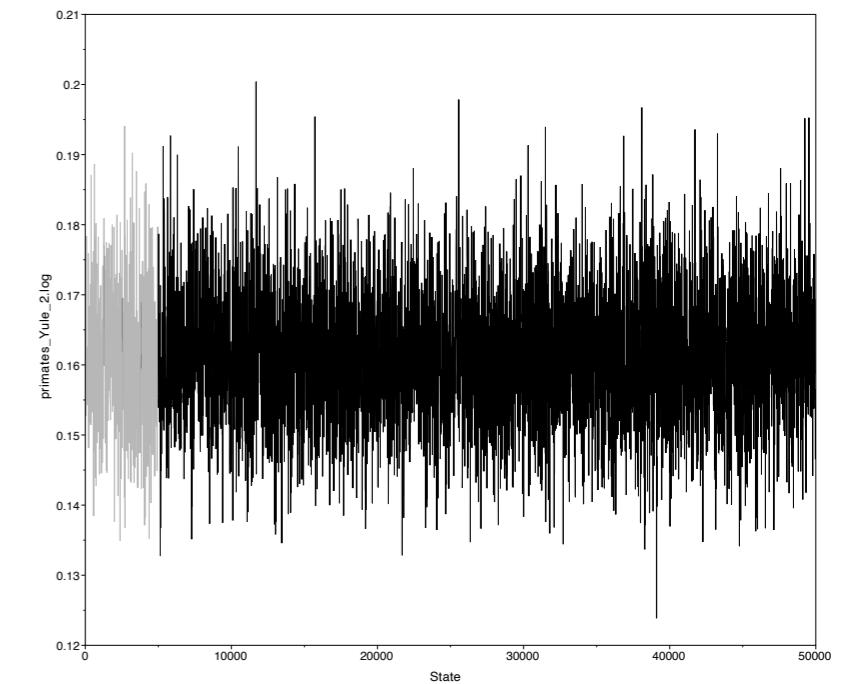
Small change
ESS: 67
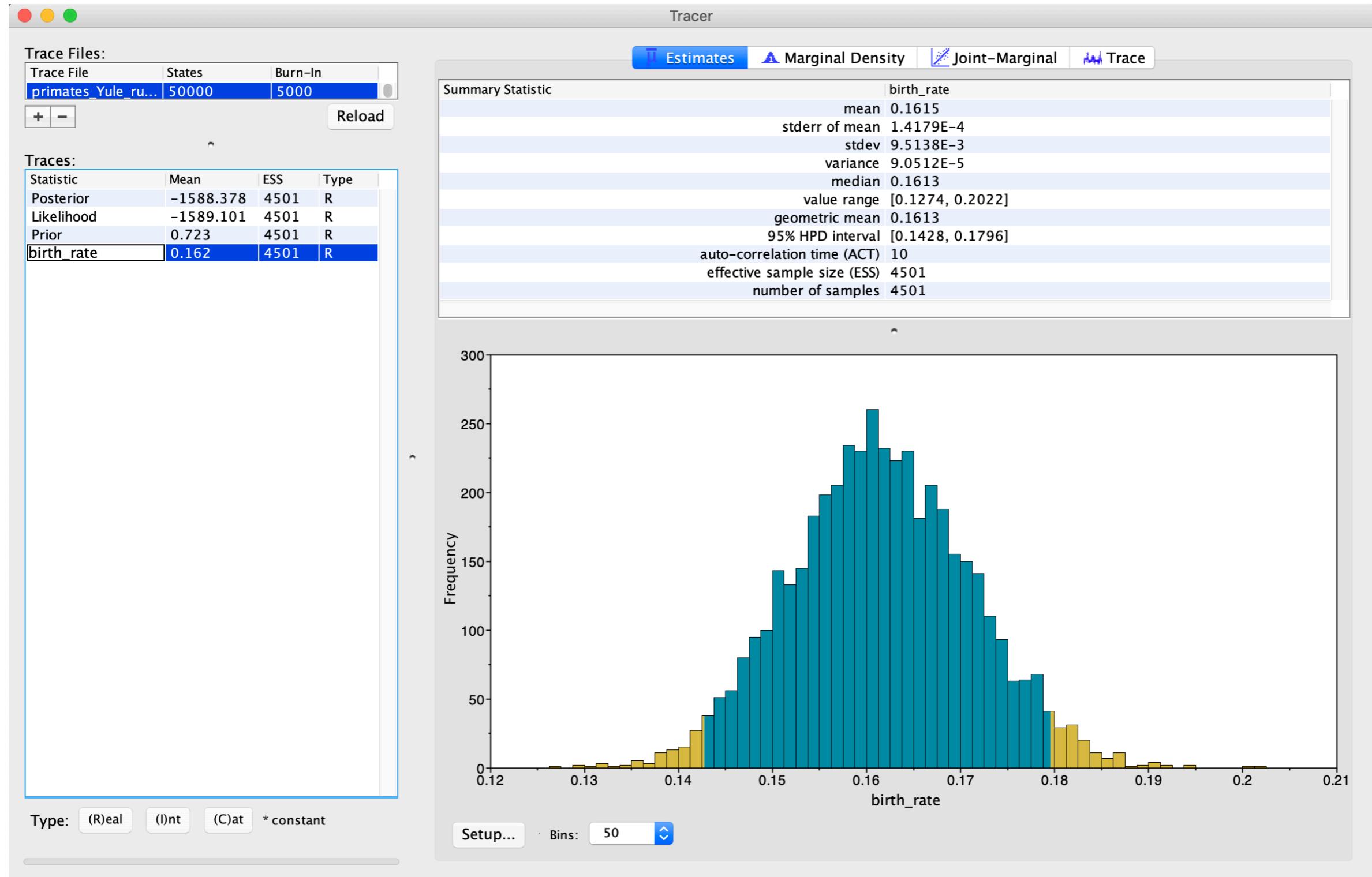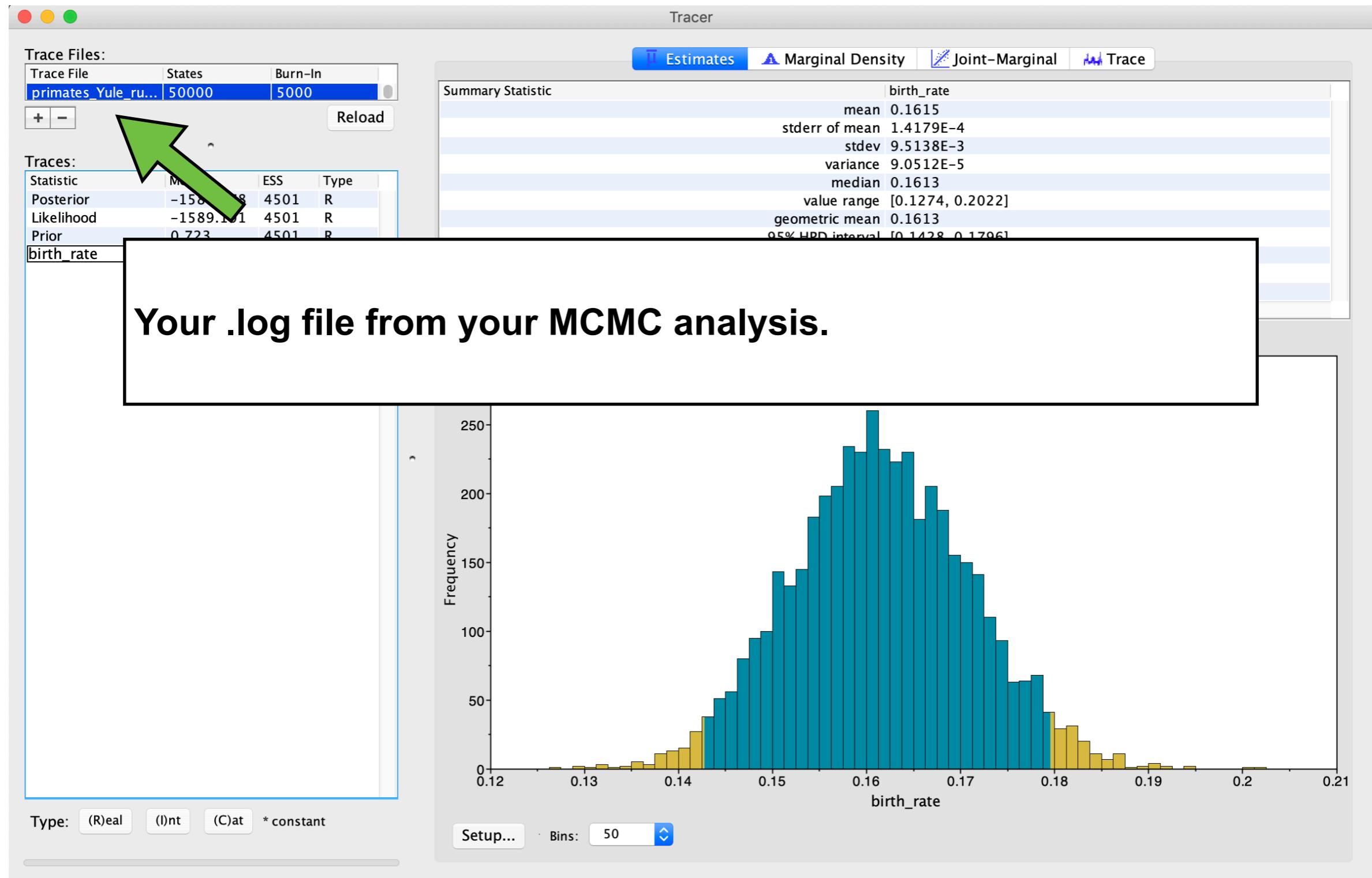
Large change
ESS: 115

Medium change
ESS: 3690

- Good acceptance rate is 0.45 for single parameter and 0.23 for multiple parameters
- Auto-tuning can set the tuning parameter to achieve good acceptance rates
- Effective sample size (ESS) should be >625

Summarizing MCMC output

Summarizing MCMC output


Your .log file from your MCMC analysis.

Summarizing MCMC output

Summarizing MCMC output

Several important summaries for your parameter:

  Posterior mean

  95% HPD (Highest Posterior Density)

  Effective Sample Size

Summarizing MCMC output

Tracer

Trace Fi[le]
primate[s]

Traces:
Statistic
Posterio[r]

Posterior distribution (95% credible intervals is colored).

Likelihood	−1589.101	4501	R
Prior	0.723	4501	R
birth_rate	0.162	4501	R

geometric mean 0.1613
95% HPD interval [0.1428, 0.1796]
auto-correlation time (ACT) 10
effective sample size (ESS) 4501
number of samples 4501
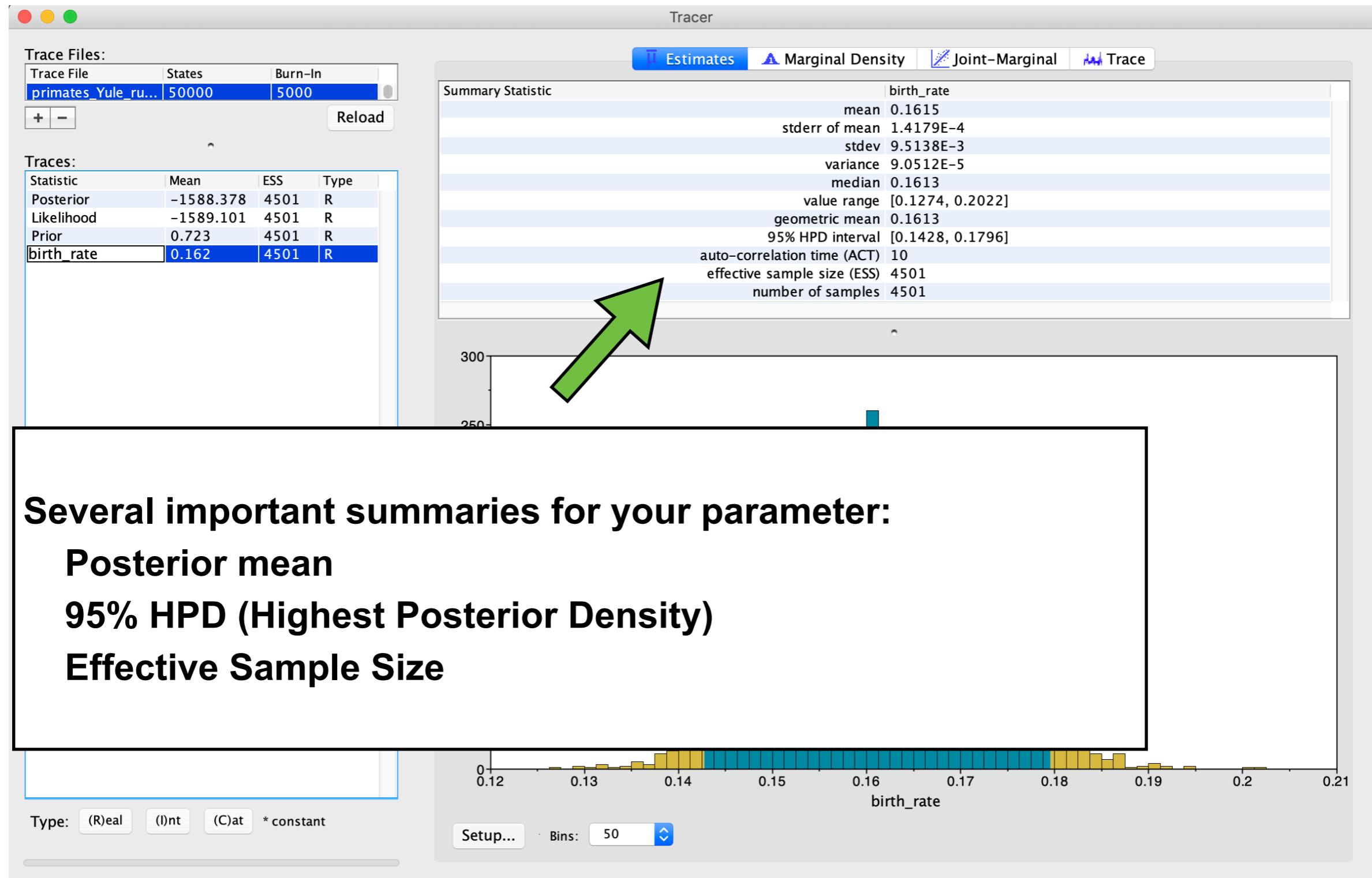
Type: (R)eal (I)nt (C)at * constant
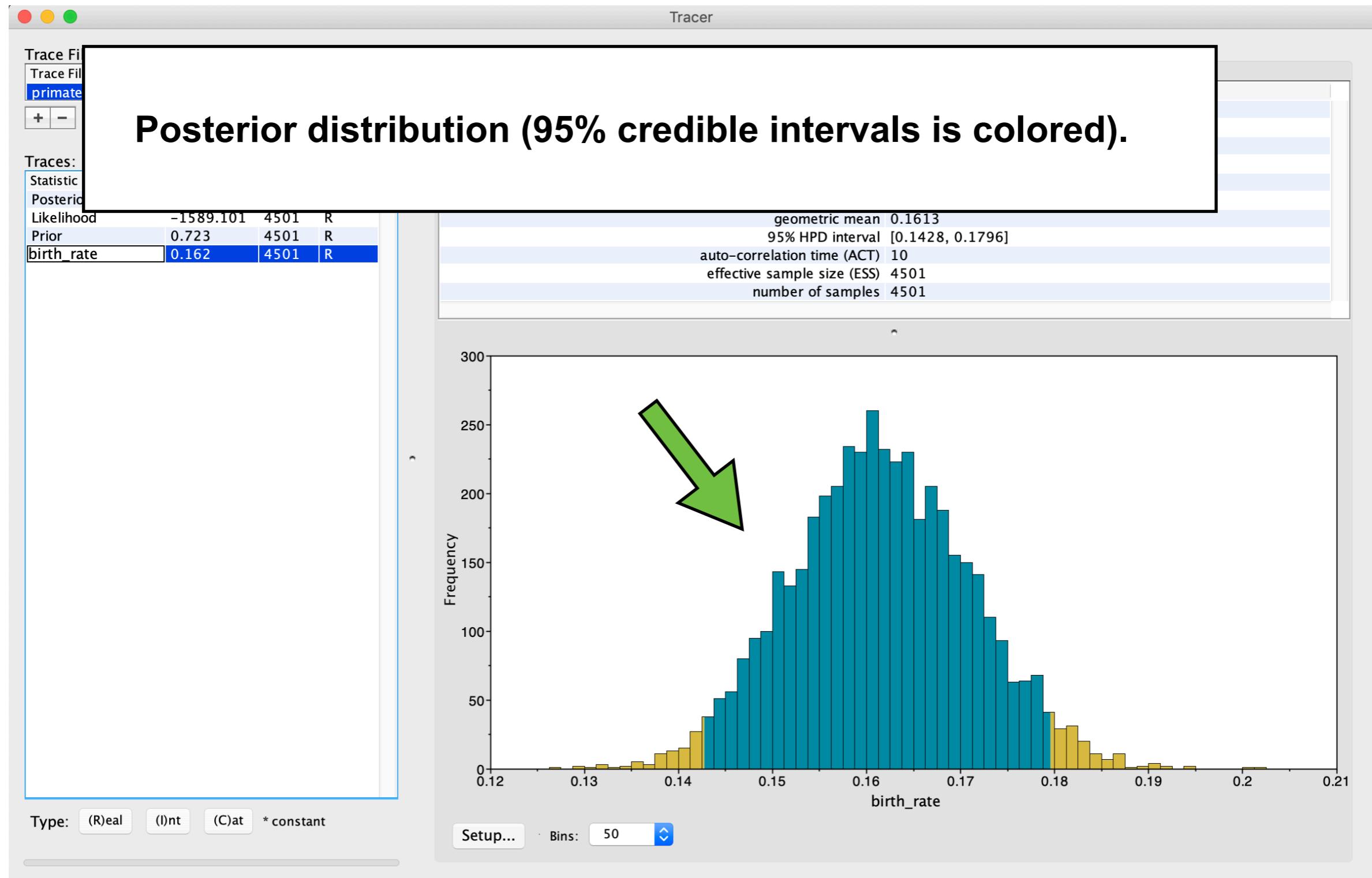
Setup... · Bins: 50

birth_rate

Assessing MCMC convergence

1. Precision

The uncertainty of the estimator must be smaller than a given tolerance value. That is, longer chains or more samples will not lead to significantly different estimates, given the tolerated uncertainty.

2. Reproducibility

Repeated chains from random starting values will give the same estimates, given the tolerated uncertainty.

Assessing MCMC convergence
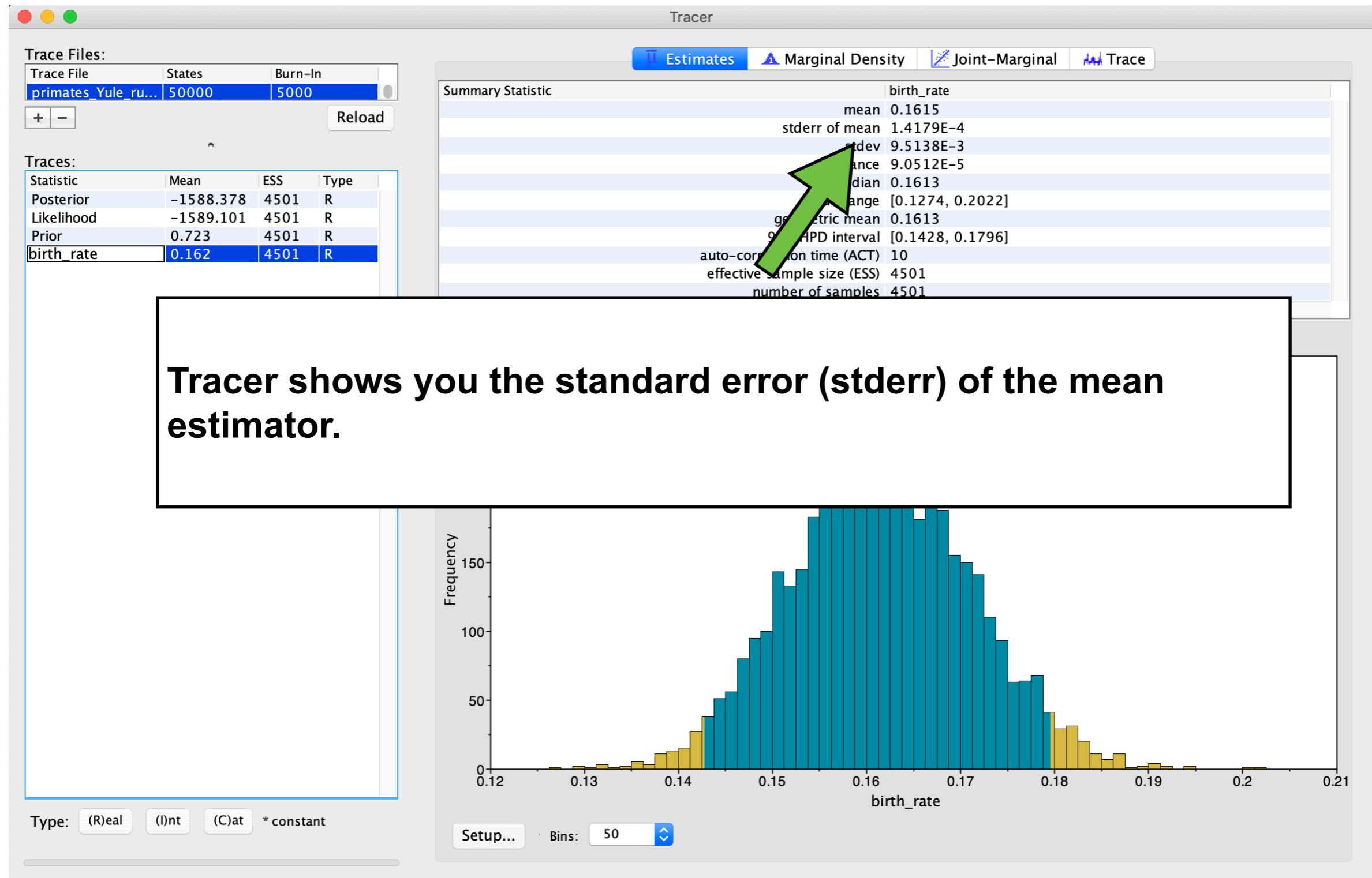
1. Precision

The uncertainty of the estimator must be smaller than a given tolerance value. That is, longer chains or more samples will not lead to significantly different estimates, given the tolerated uncertainty.

2. Reproducibility
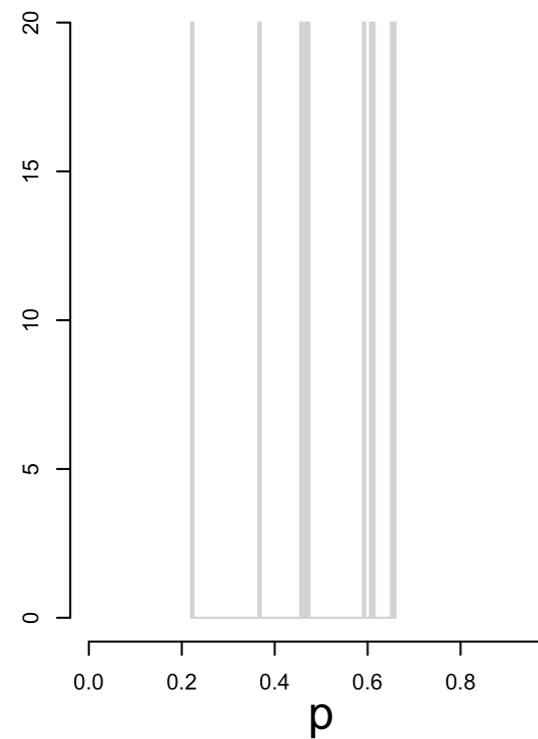
Repeated chains from random starting values will give the same estimates, given the tolerated uncertainty.

Precision of an estimator
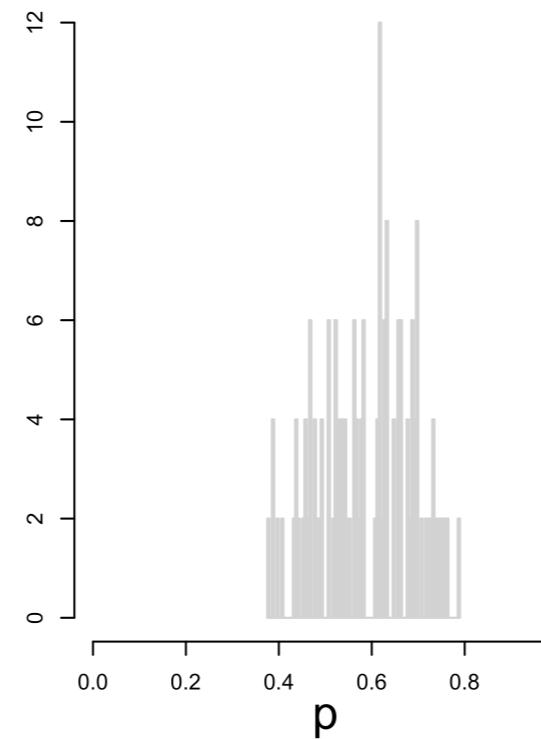
Tracer shows you the standard error (stderr) of the mean estimator.

Precision of an estimator
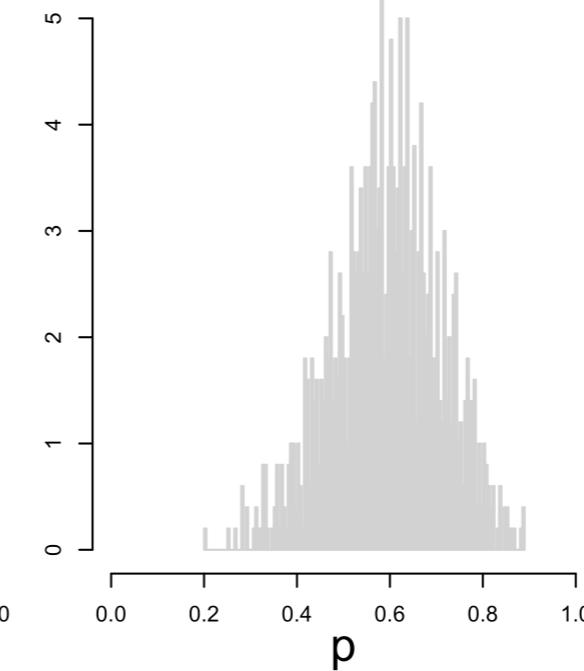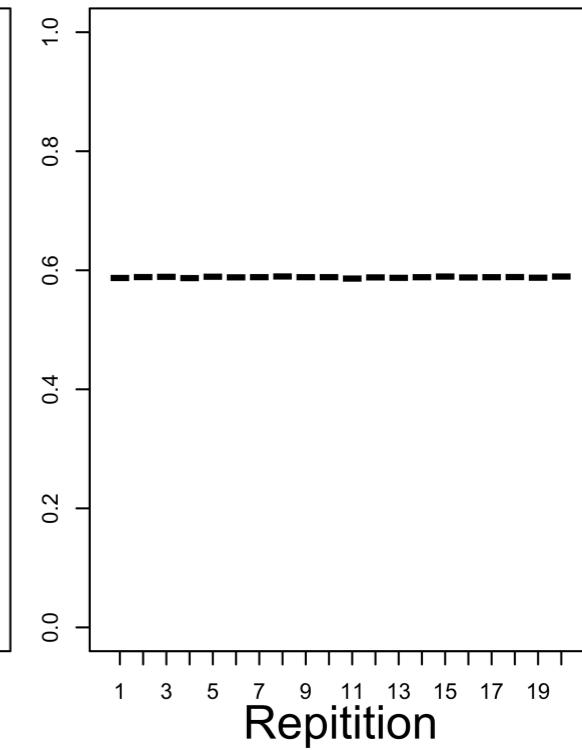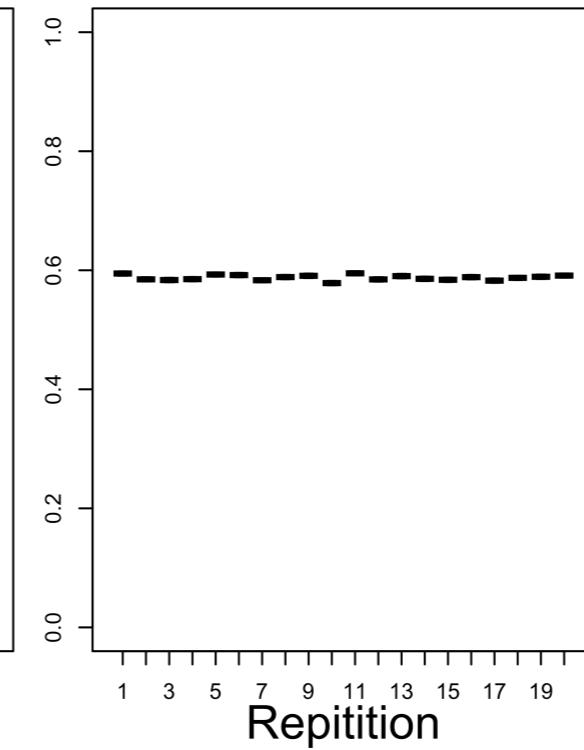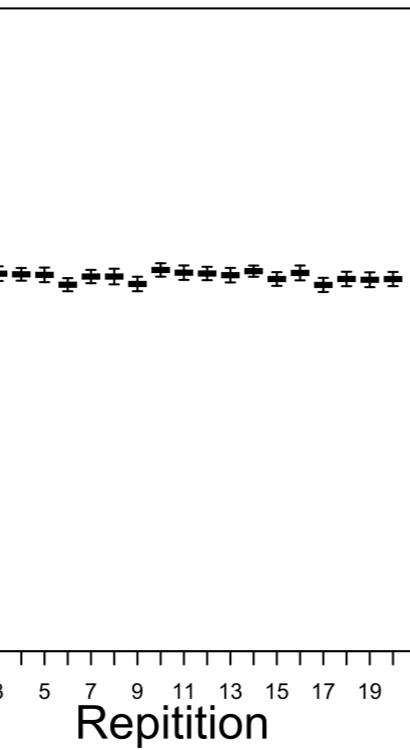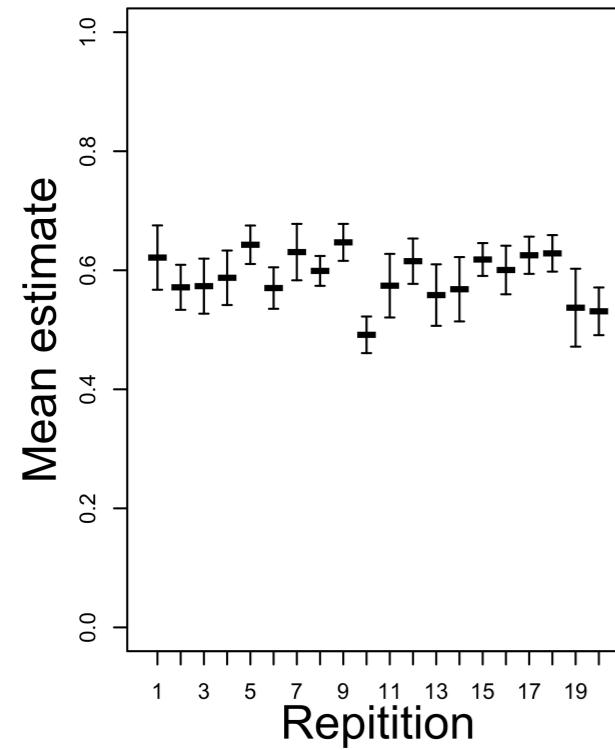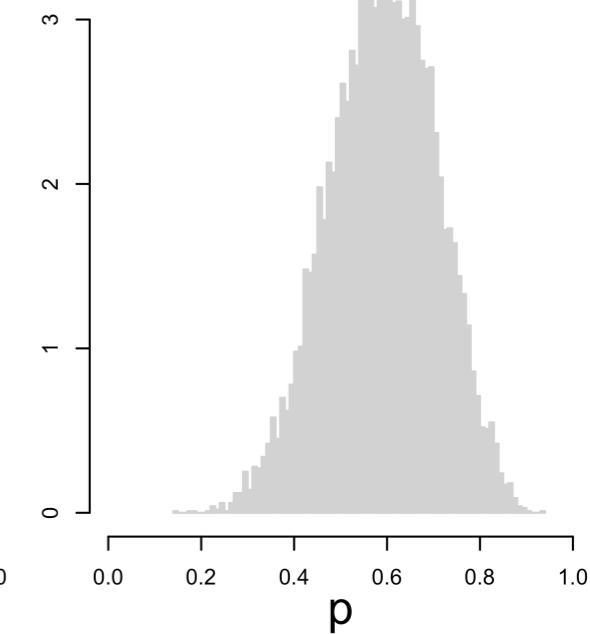
Assessing MCMC convergence

1. Precision
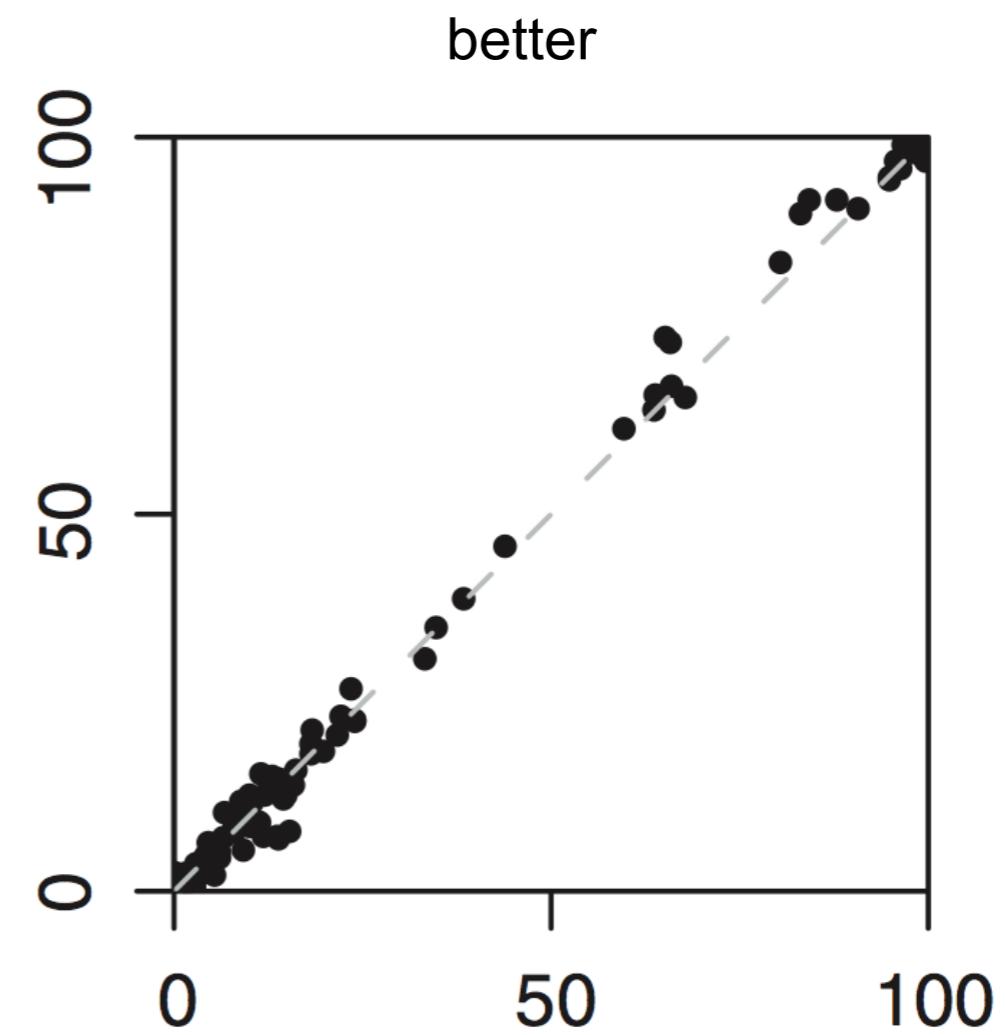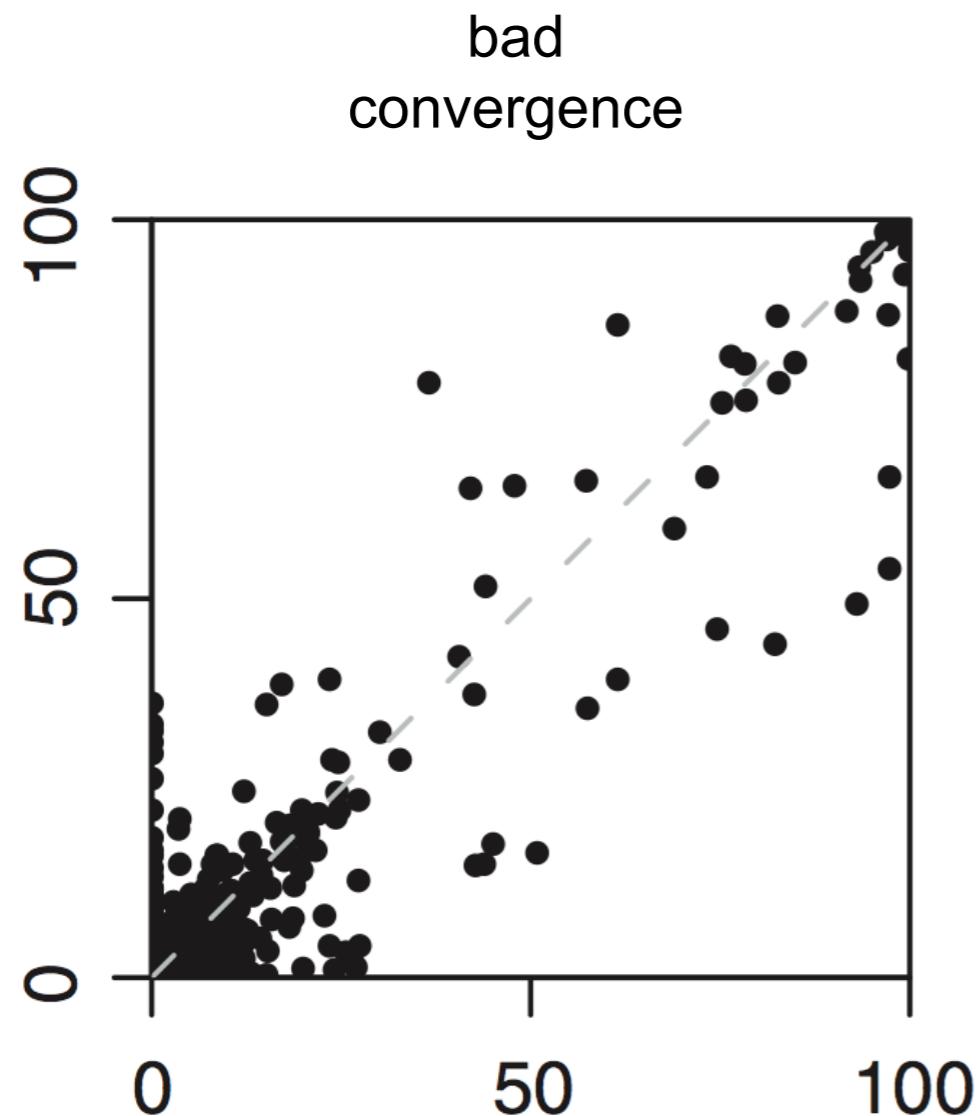
The uncertainty of the estimator must be smaller than a given tolerance value. That is, longer chains or more samples will not lead to significantly different estimates, given the tolerated uncertainty.

2. Reproducibility

Repeated chains from random starting values will give the same estimates, given the tolerated uncertainty.

Comparing trees from replicates

Example: plot of posterior probabilities of clades sampled by two runs

Improving MCMC Performance

If the chain has not converged (according to precision and/or reducibility), what can you do?

1. Burnin

Are all samples drawn from the stationary distribution?

2. Mixing

Is the chain efficiently integrating over the joint posterior probability?

3. Sample Size

Have we collected enough samples to adequately describe the posterior probability distribution?

Improving MCMC Performance

If the chain has not converged (according to precision and/or reducibility), what can you do?

1. Burnin

Are all samples drawn from the stationary distribution?

2. Mixing

Is the chain efficiently integrating over the joint posterior probability?

3. Sample Size

Have we collected enough samples to adequately describe the posterior probability distribution?

Burnin
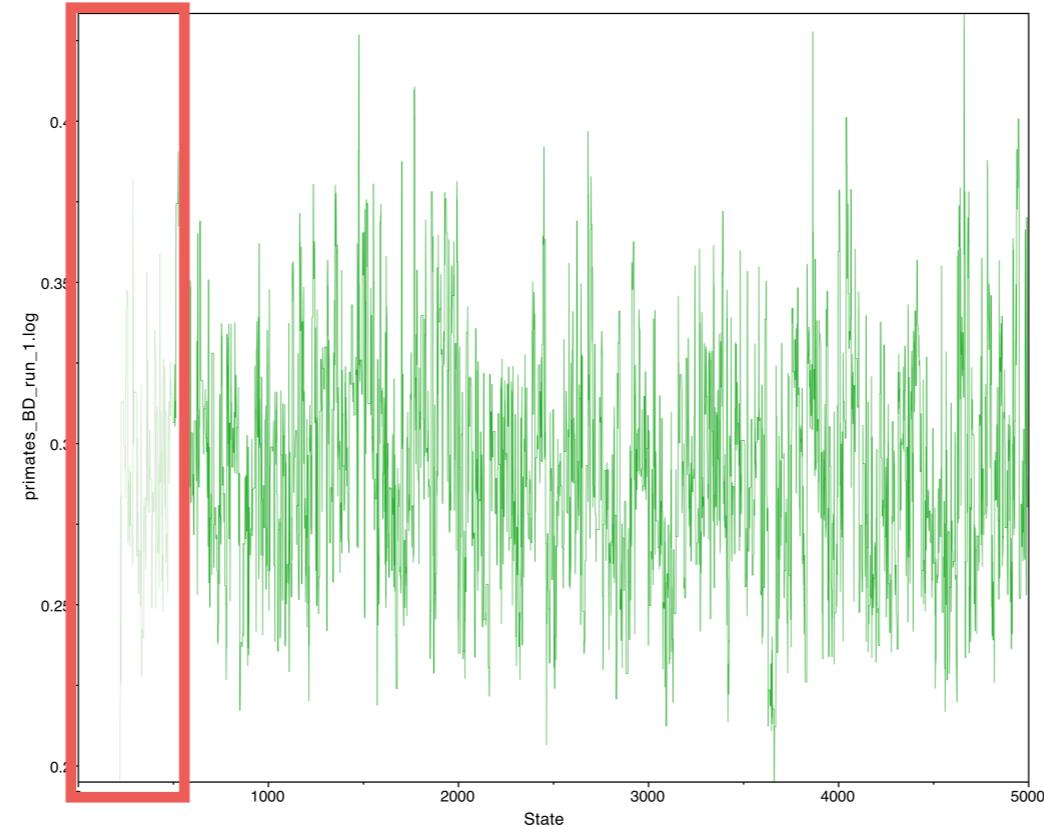
mymcmc = mcmc(mymodel, moves, monitors, nruns=2, combine="mixed")

mymcmc.burnin(generations=2000,tuningInterval=200)
mymcmc.run(generations=20000,tuningInterval=200)

You can run a pre-burnin in RevBayes.

Or you can afterwards remove the first 10% or 25% as burnin.

Improving MCMC Performance

If the chain has not converged (according to precision and/or reducibility), what can you do?

1. Burnin

Are all samples drawn from the stationary distribution?
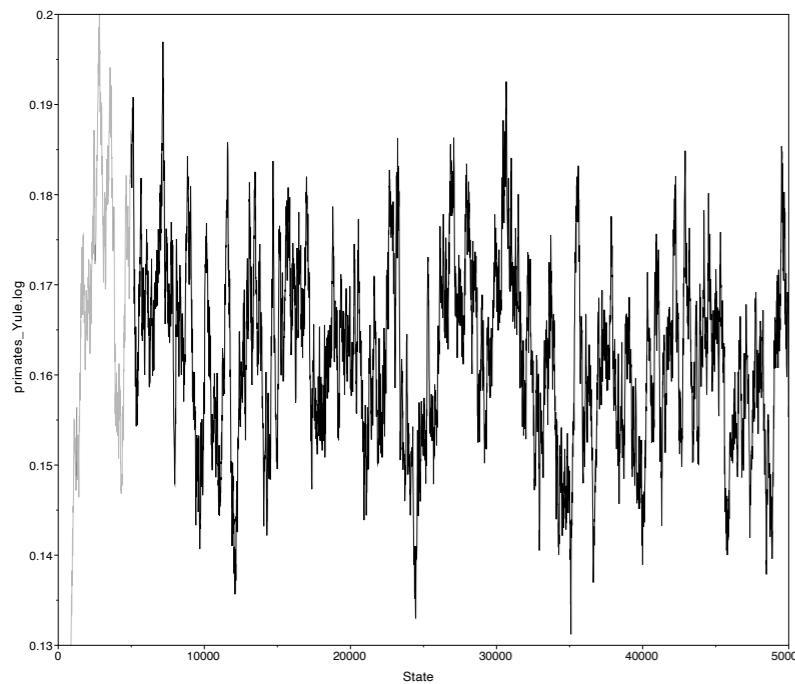
2. Mixing

Is the chain efficiently integrating over the joint posterior probability?

3. Sample Size
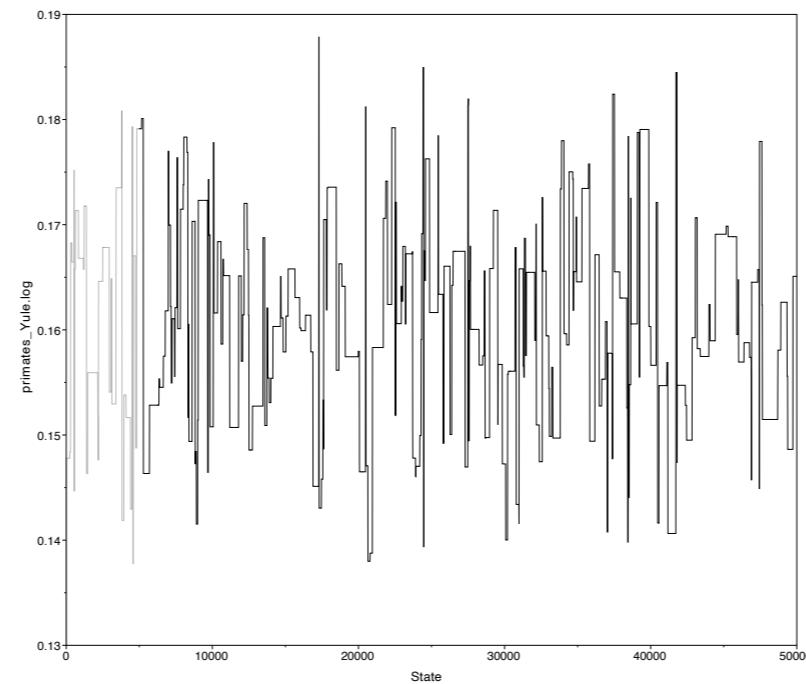
Have we collected enough samples to adequately describe the posterior probability distribution?

Mixing
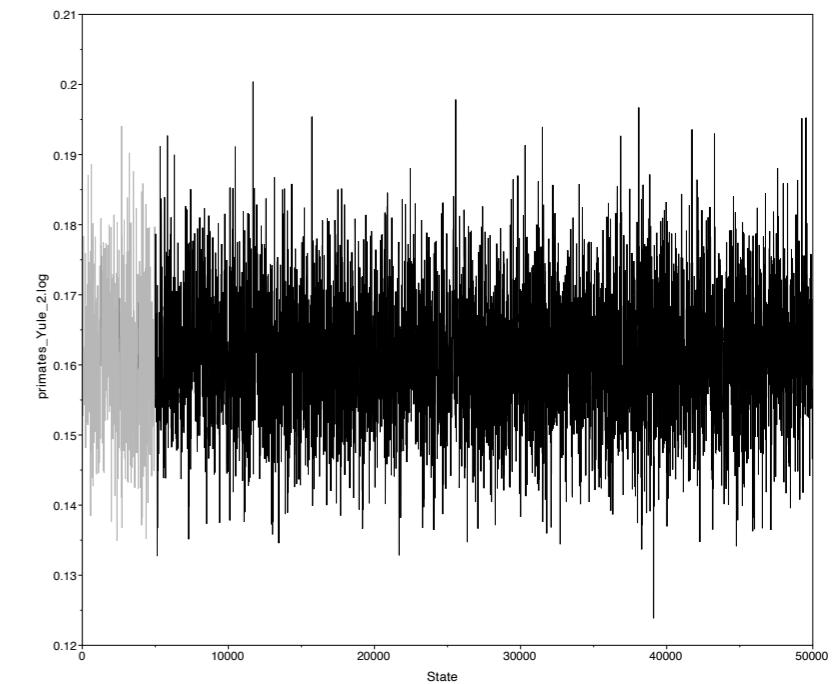
Small change
ESS: 67

Large change
ESS: 115

Medium change
ESS: 3690

- Good acceptance rate is 0.45 for single parameter and 0.23 for multiple parameters
- Auto-tuning can set the tuning parameter to achieve good acceptance rates
- Effective sample size (ESS) should be >625

Mixing in RevBayes

mymcmc = mcmc(mymodel, moves, monitors, nruns=2, combine="mixed")
mymcmc.run(generations=20000,tuningInterval=200)

mymcmc.operatorSummary()

Improving MCMC Performance

If the chain has not converged (according to precision and/or reducibility), what can you do?

1. Burnin

Are all samples drawn from the stationary distribution?

2. Mixing

Is the chain efficiently integrating over the joint posterior probability?

3. Sample Size

Have we collected enough samples to adequately describe the posterior probability distribution?

More about MCMC convergence assessment:

- See talk by Luiza Fabreti
- See the R package convenience (https://github.com/lfabreti/convenience)
- https://revbayes.github.io/tutorials/convergence/