MCMC

Numerical methods for Bayes

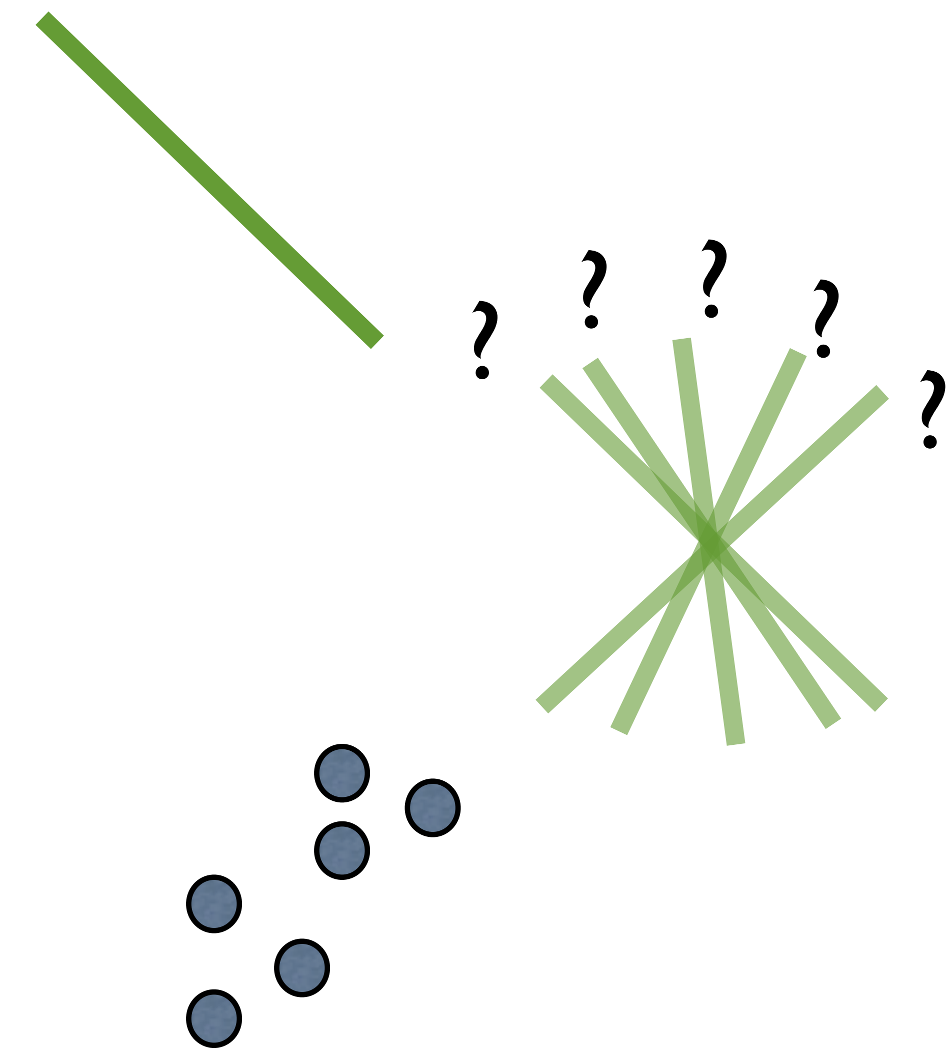

SPOILER
01
HANG THE DJ
ALERT
NETFLIX
BLACK MIRROR


1000
SIMULATIONS COMPLETED
998
REBELLIONS DETECTED
99.8%
MATCH

What are the chances that a solitaire laid out with 52 cards will come out successfully?

Stanislaw Ulam, 1946, LANL

Goal: Find the laws that govern the functioning and the interactions among nature's living organisms.



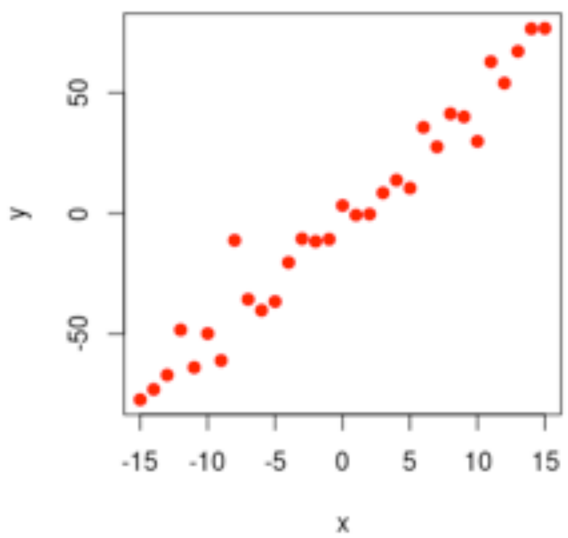
How? Through a model, $M(\theta)$

What do we know? $P(\theta)$

What do we have? Some data (Y)

What do we want to know? $P(\theta | Y) = \dfrac{P(Y|\theta)\, P(\theta)}{P(Y)}$

Rejection sampling



Model:

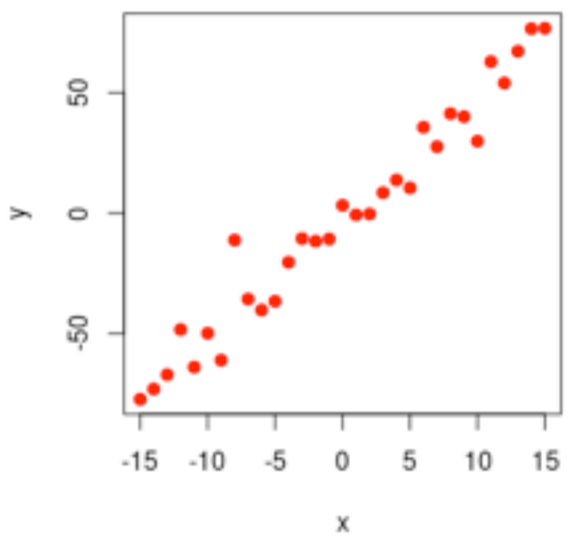
y = b + a*x

Priors:

$P(a) \sim U[a_{min}, a_{max}]$

$P(b) \sim U[b_{min}, b_{max}]$

Rejection sampling



Model:

y = b + a*x + Ɛ

N(0, sd)

y ~ N(b + a*x, sd)

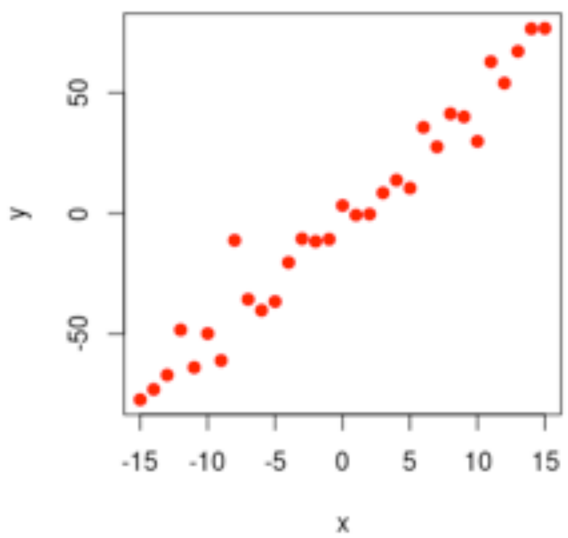
Priors:

P(a) ~ U[a_{min}, a_{max}]

P(b) ~ U[b_{min}, b_{max}]

P(sd) ~ U[sd_{min}, sd_{max}]

Rejection sampling



Model:

y ~ N(a*x, 10)

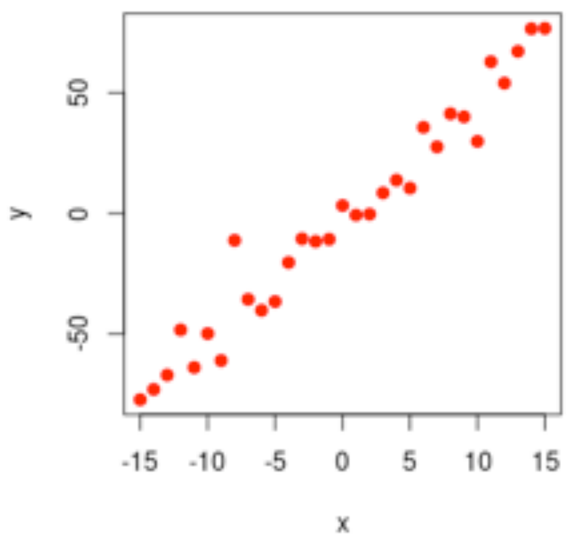
Priors:

P(a) ~ U[a_{min}, a_{max}]

b = 0

sd =10

Rejection sampling



1.draw random values from prior

Model:

y ~ N(a*x, 10)

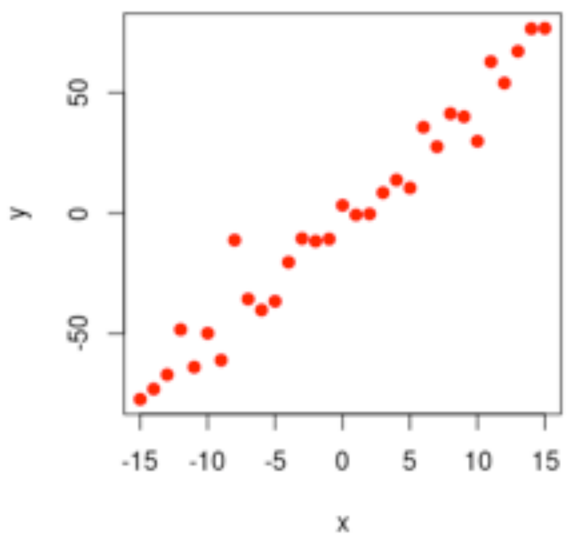
Priors:

$P(a) \sim U[a_{min}, a_{max}]$ ⟶ a_1

b = 0

sd =10

Rejection sampling



1. draw random values from priors
2. calculate likelihood

i.e. the probability of observing data y given the model and drawn parameter values:

$y \ ?\!\sim N(a_1 * x, 10)$

Model:

$y \sim N(a*x, 10)$

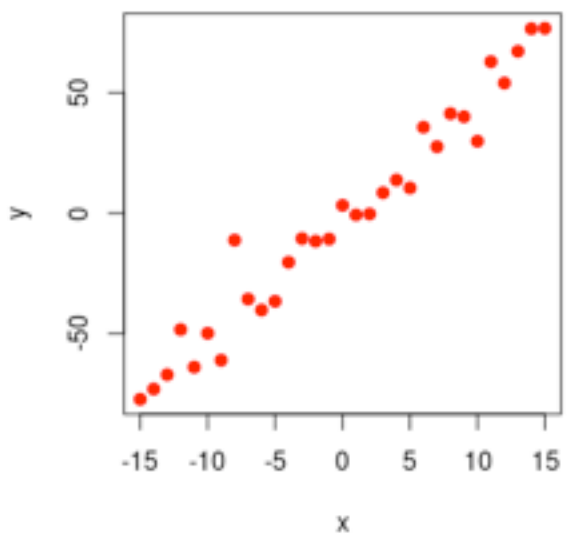
Priors:

$P(a) \sim U[a_{min}, a_{max}]$ ⟶ a_1

b = 0

sd = 10

Rejection sampling



1. draw random values from priors
2. calculate likelihood

i.e. the probability of observing data y given the model and drawn parameter values:

Model:

y ~ N(a*x, 10)

Priors:

P(a) ~ U[a_{min}, a_{max}] → a_1

b = 0

sd =10

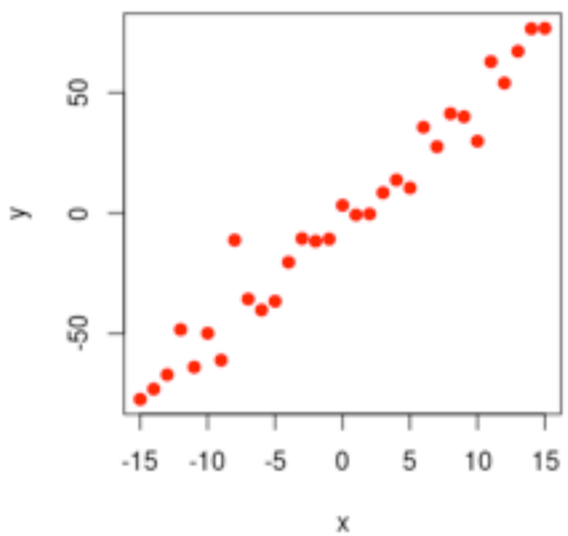
y ?~ N(a_1*x, 10)

probability density function for normal distribution

$$\frac{1}{sd\sqrt{2\pi}} e^{-\frac{(y-(b+ax))^2}{2sd^2}}$$

in R: `sum(dnorm(y, mean = 0+ a1*x, sd = 10))`

Rejection sampling



1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

Model:

$y \sim N(a*x, 10)$

$$P(\theta|Y) = \frac{P(Y|\theta)\, P(\theta)}{P(Y)}$$

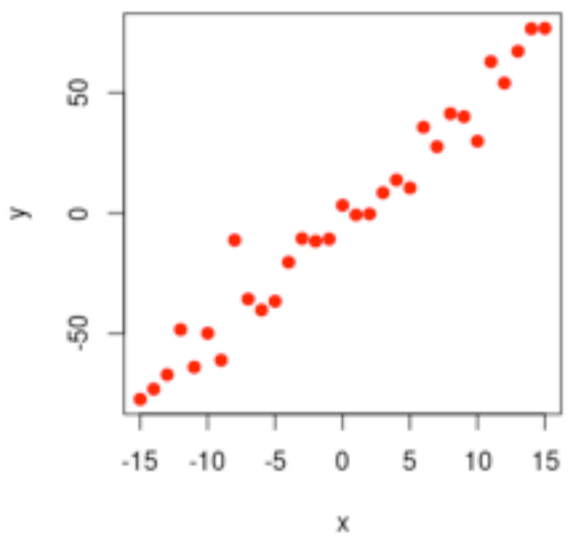
Priors:

$P(a) \sim U[a_{min}, a_{max}] \longrightarrow a_1$

$b = 0$

$sd = 10$

Rejection sampling



1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

Model:

$y \sim N(a*x, 10)$

$$P(a|y) \propto P(y|a)\,P(a)$$

unnormalized posterior PDF

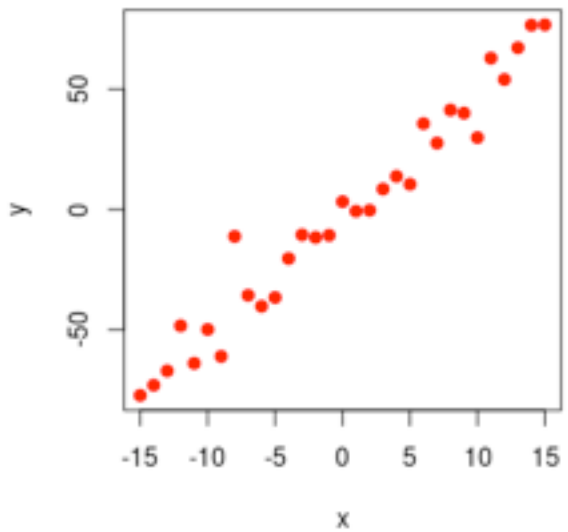
Priors:

$P(a) \sim U[a_{min}, a_{max}]$

$b = 0$

$sd = 10$

Rejection sampling



~~1.draw random values from priors~~

~~2. calculate likelihood~~

~~3. accept value proportional to likelihood~~

a. Sample many values from prior

```
Asamples <- runif(1000, amin, amax)
```

Model:

y ~ N(a*x, 10)

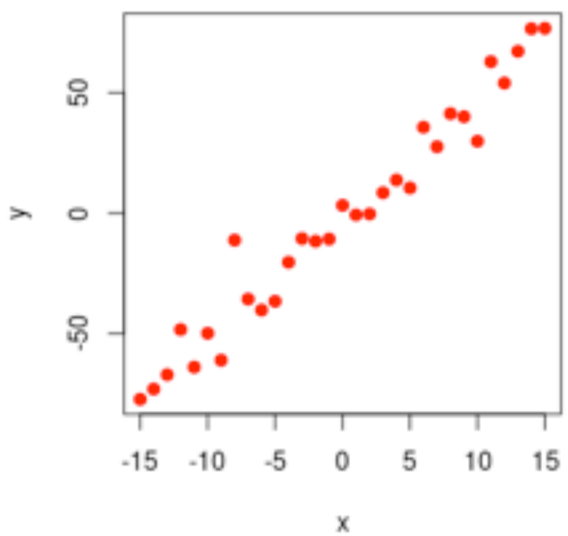
Priors:

P(a) ~ U[a_{min}, a_{max}]

b = 0

sd =10

Rejection sampling



~~1.draw random values from priors~~
~~2. calculate likelihood~~
~~3. accept value proportional to likelihood~~

a. Sample many values from prior

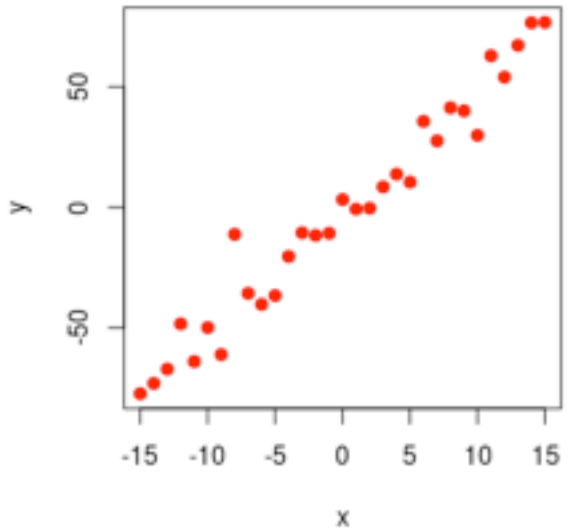
b. Calculate unnormalized posterior values

Model:

y ~ N(a*x, 10)

```
for(i in 1:1000)
LL          <- sum(dnorm(y, mean = Asamples[i]*x, sd =10))
prior       <- dunif(Asamples[i], amin, amax)
Psamples    <- LL * prior
```

Rejection sampling



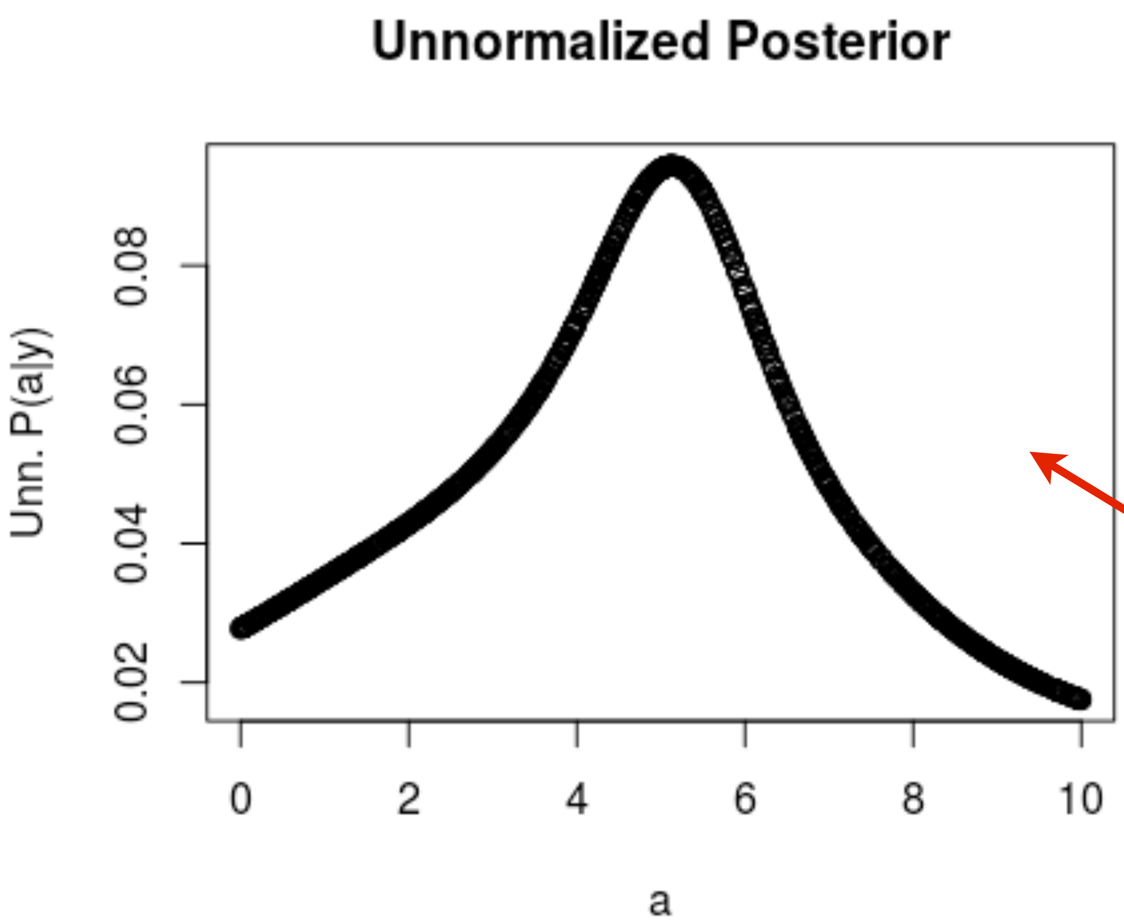
~~1.draw random values from priors~~

~~2. calculate likelihood~~

~~3. accept value proportional to likelihood~~

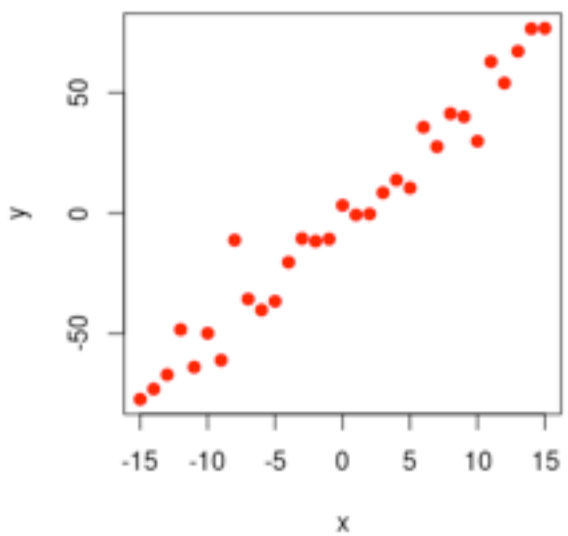
a. Sample many values from prior

b. Calculate unnormalized posterior values



This is proportional to what I want to sample from

Rejection sampling



~~1.draw random values from priors~~

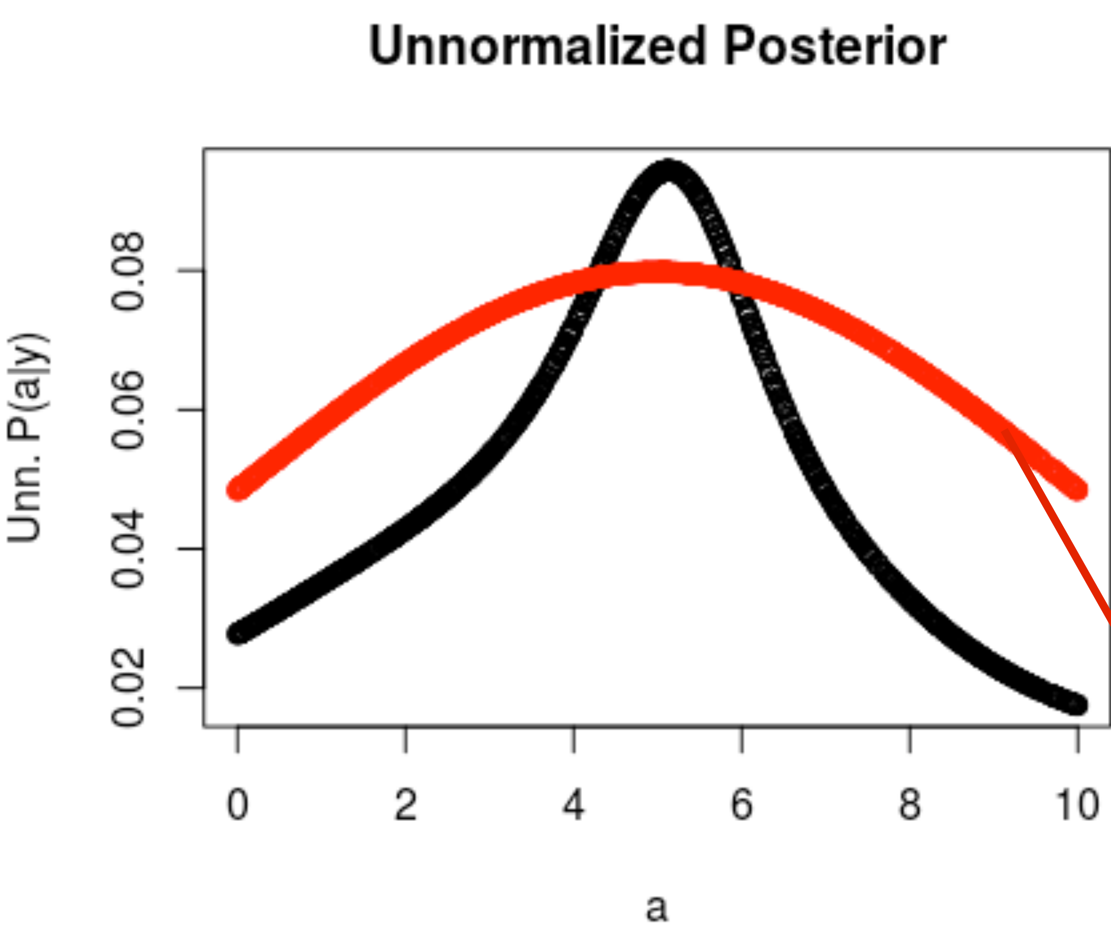
~~2. calculate likelihood~~

~~3. accept value proportional to likelihood~~

a. Sample many values from prior

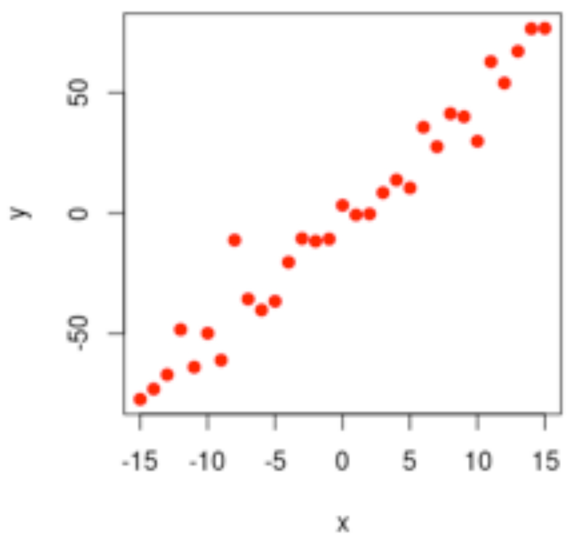
b. Calculate unnormalized posterior values

c. Choose a distribution from which sampling is easy

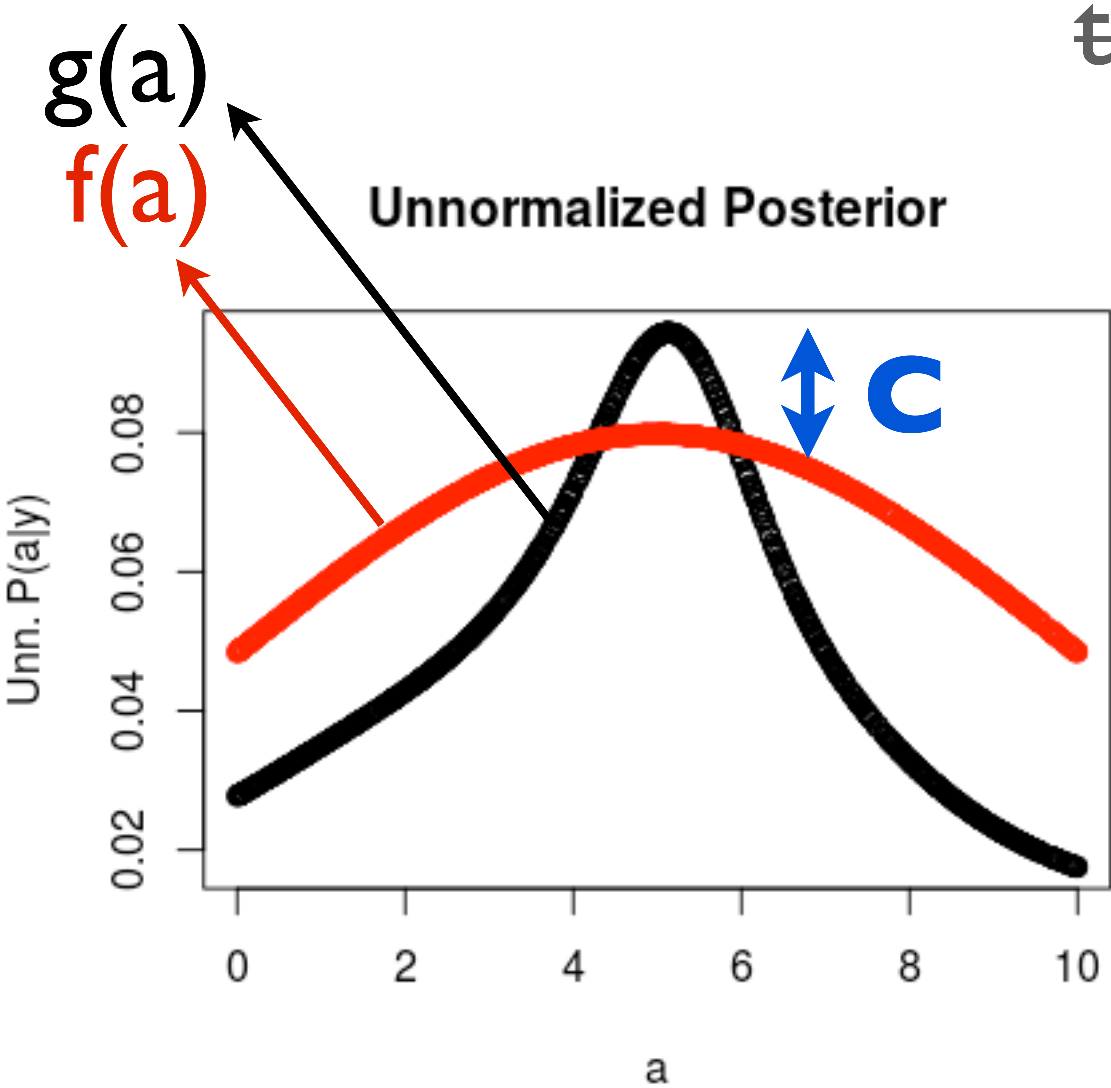


N(5,5) : proposal or envelope distribution

Rejection sampling



~~1.draw random values from priors~~
~~2. calculate likelihood~~
~~3. accept value proportional to likelihood~~



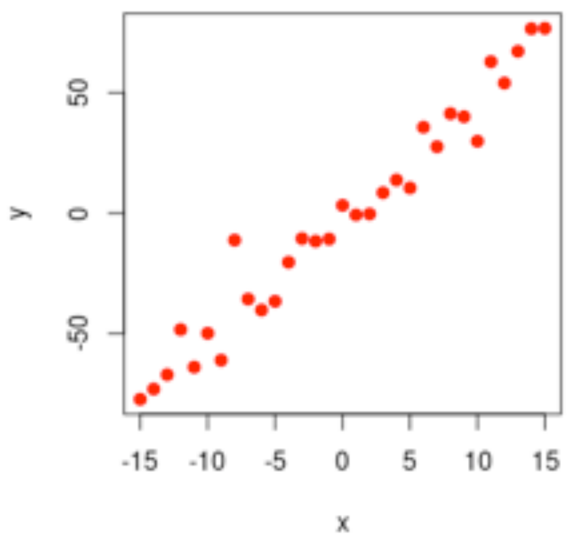
a. Sample many values from prior

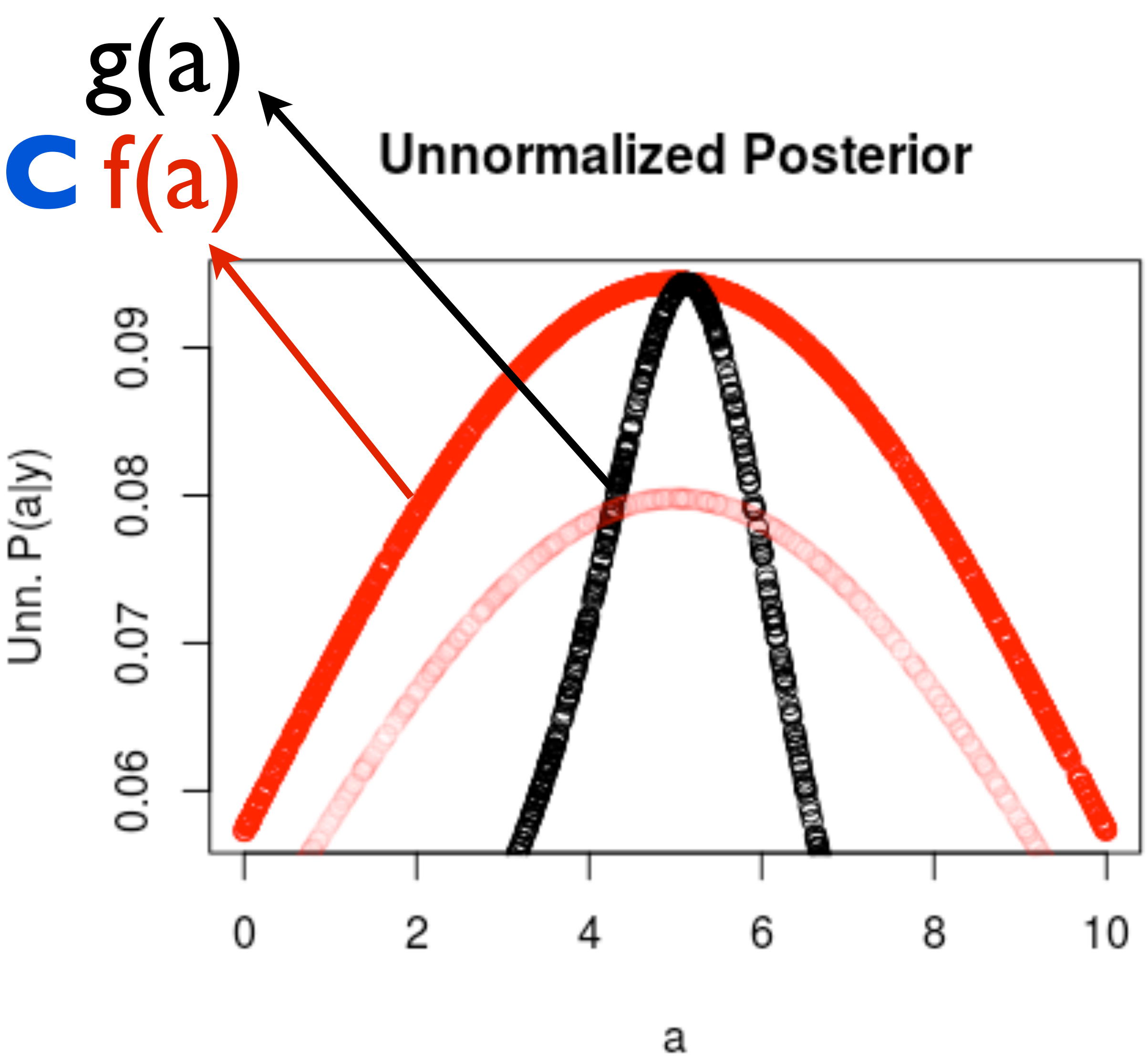
b. Calculate unnormalized posterior values

c. Choose a distribution from which sampling is easy

d. Determine scaling constant : **C** = max(g(a)/f(a))

Rejection sampling

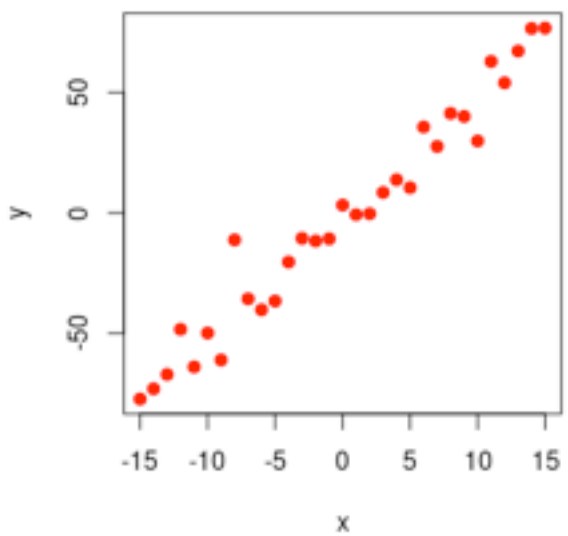




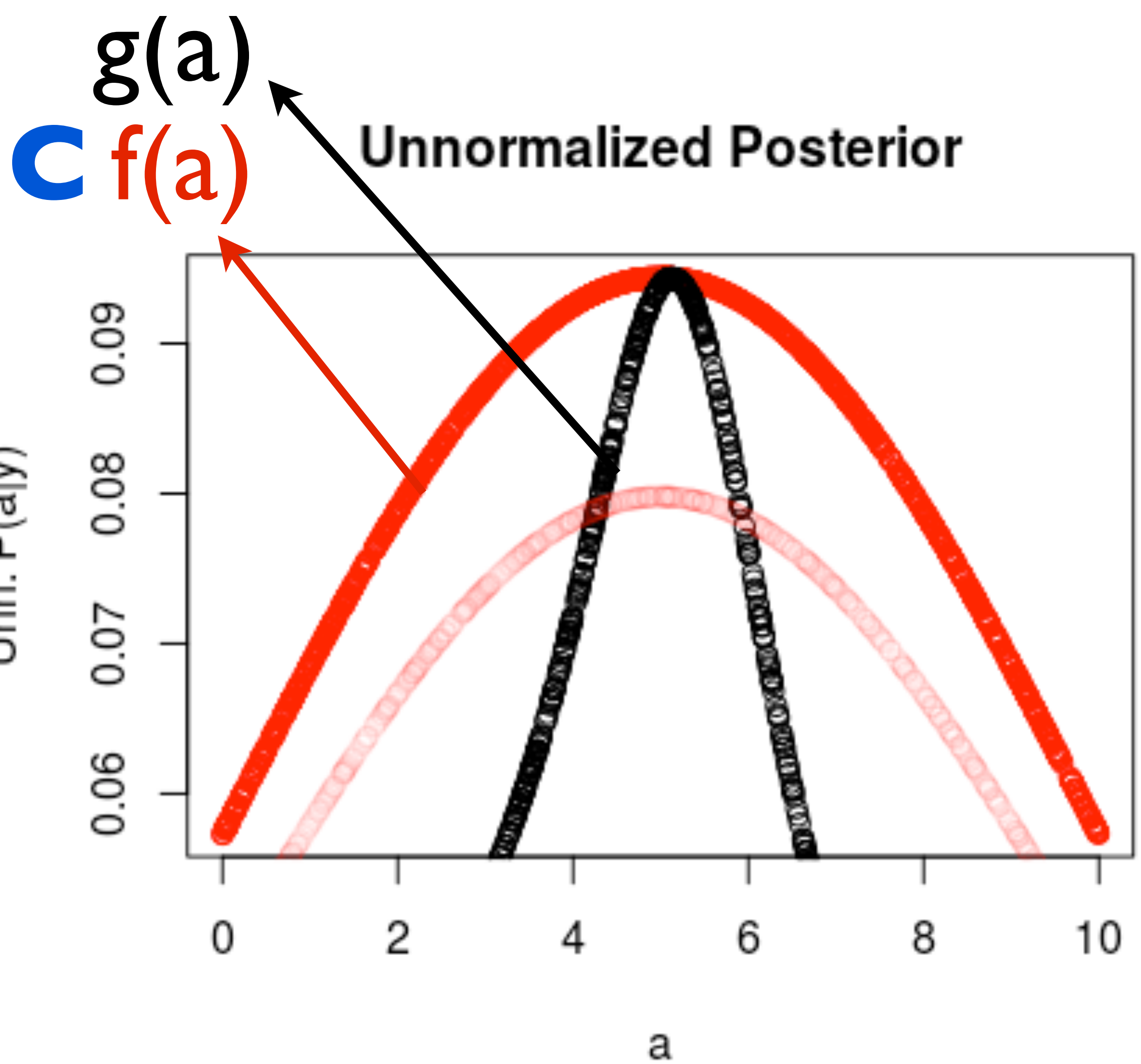
a. Sample many values from prior
b. Calculate unnormalized posterior values
c. Choose a distribution from which sampling is easy
d. Determine scaling constant : **C** = max(g(a)/f(a))
e. Scale envelope distribution

You only need to determine "C*f(a)" once!

Rejection sampling



1. draw random value from f(a)
2. calculate g(a) and f(a)
3. calculate u



$$P(a_1 | y) \propto P(y|a_1)\ P(a_1)$$

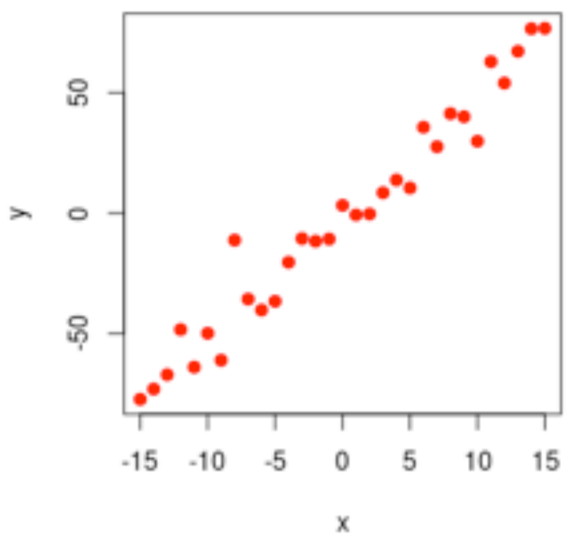
$g(a_1)=$

```
sum(dnorm(y, mean = 0+ a1*x, sd = 10))*
        dunif(a1, amin, amax)
```

$f(a_1)=$ `dnorm(a1, mean = 5, sd = 5))`

$$u = \frac{g(a_1)}{C*f(a_1)}$$

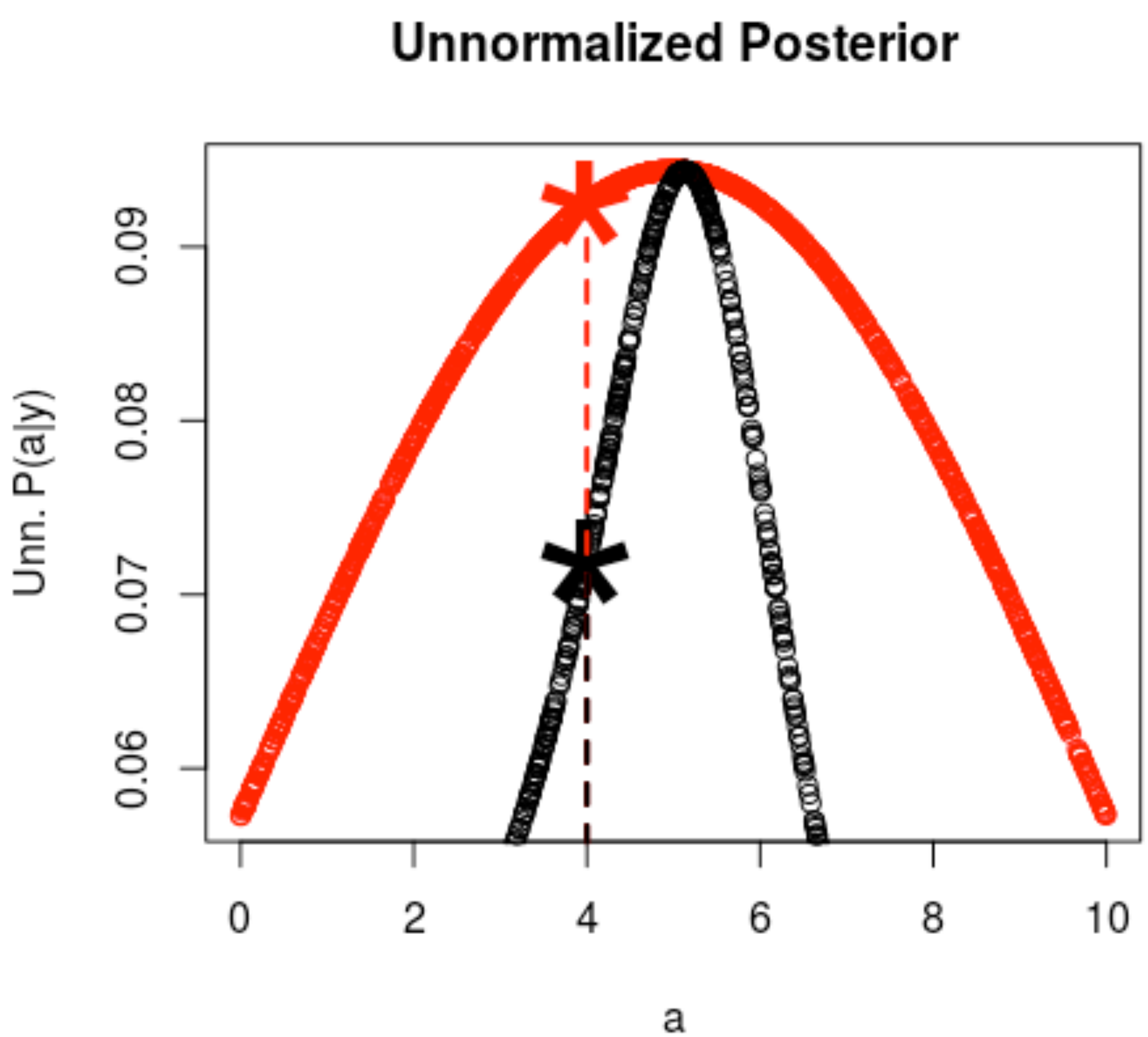
Rejection sampling



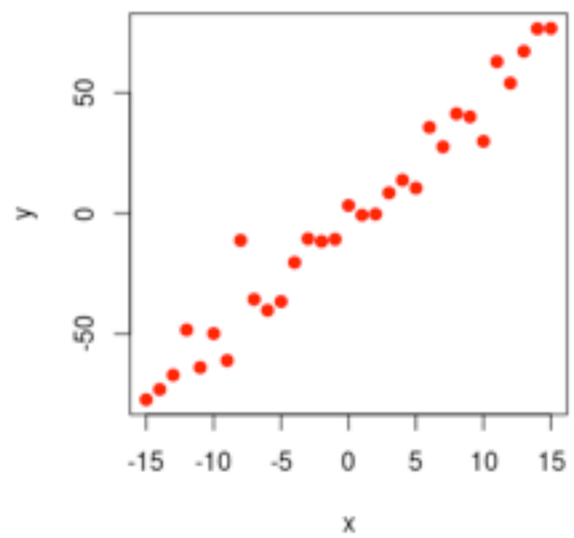
Draw random value: $a_1 = 4$

Calculate f(a) and g(a) $g(a_1), f(a_1)$

Calculate u: $u = \frac{g(a_1)}{C*f(a_1)}$



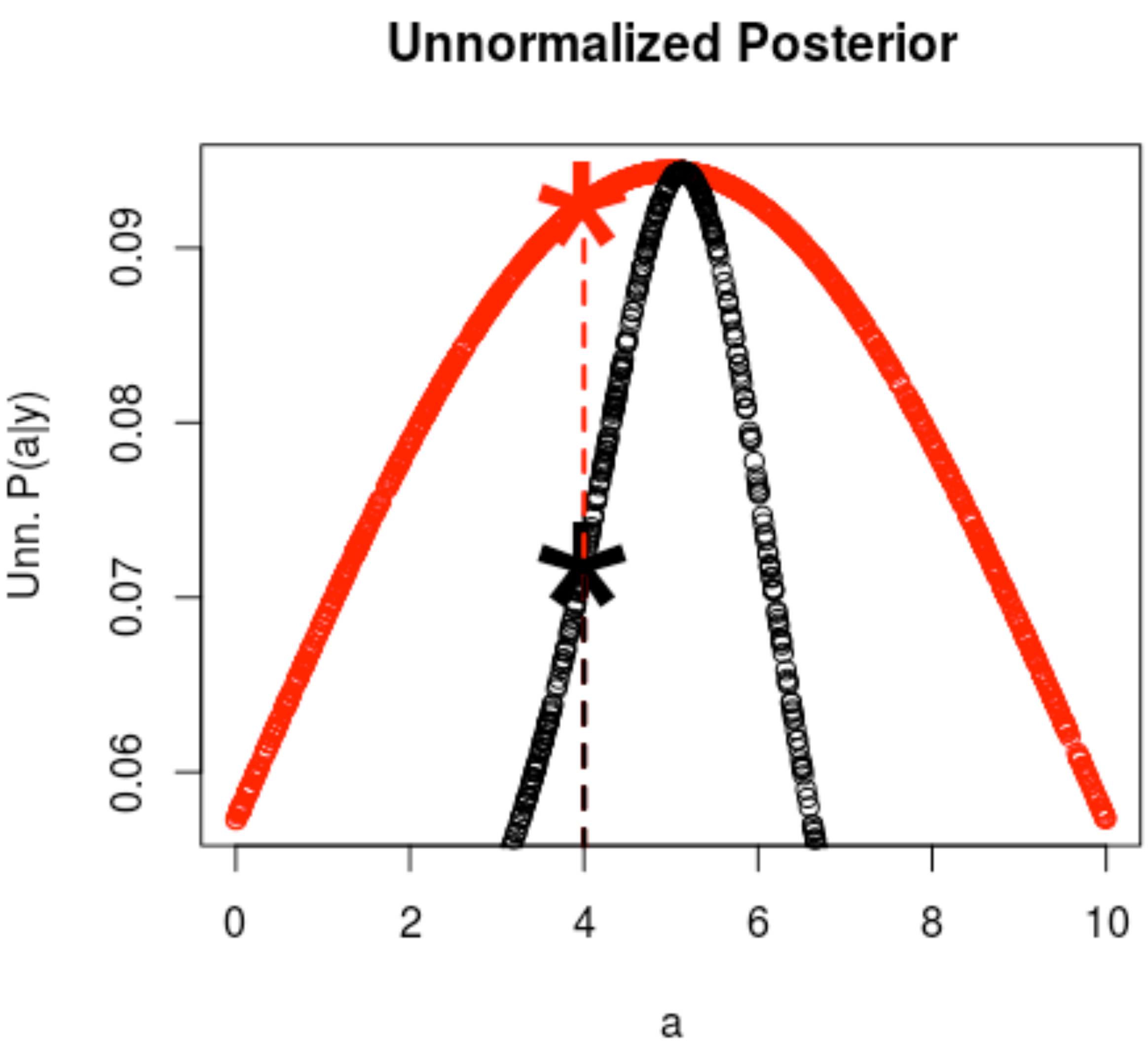
Rejection sampling



Draw random value: $a_1 = 4$

Calculate f(a) and g(a) $g(a_1), f(a_1)$

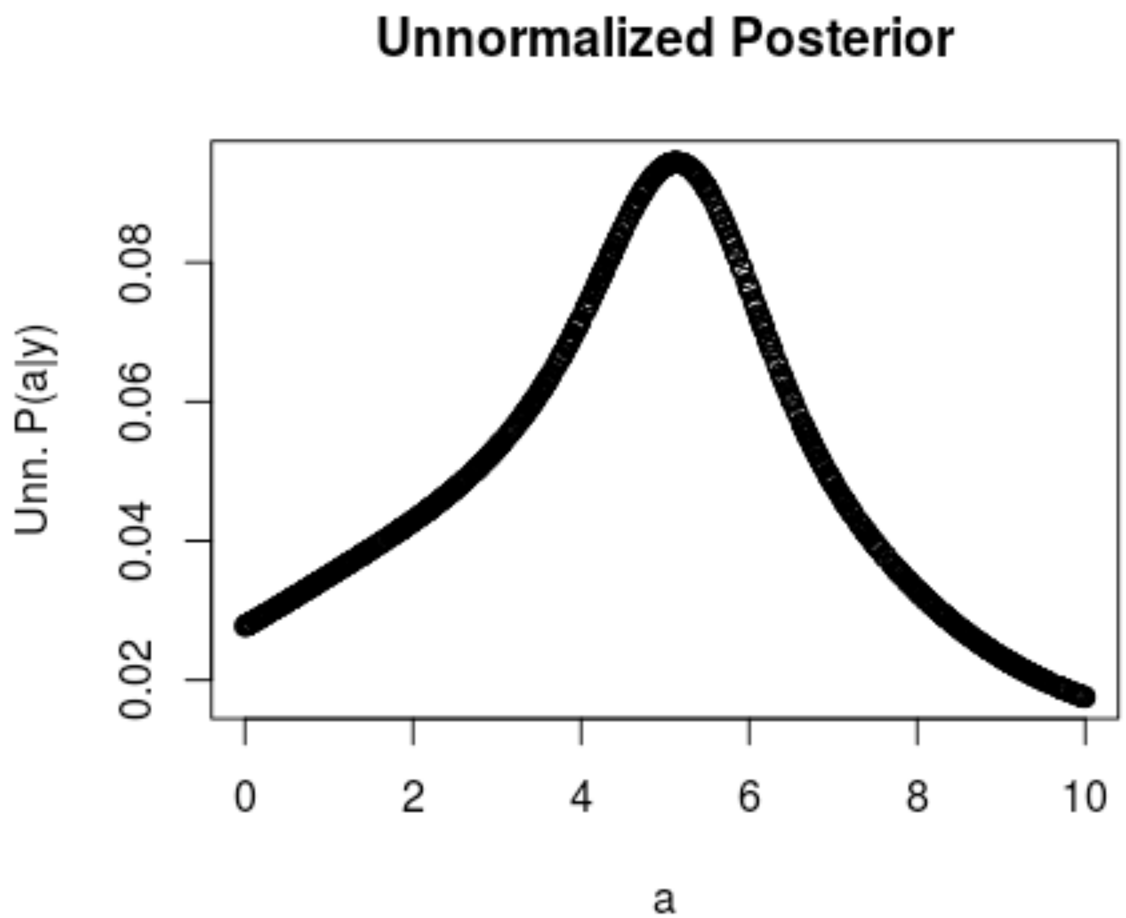
Calculate u: $u = \frac{g(a_1)}{C*f(a_1)}$

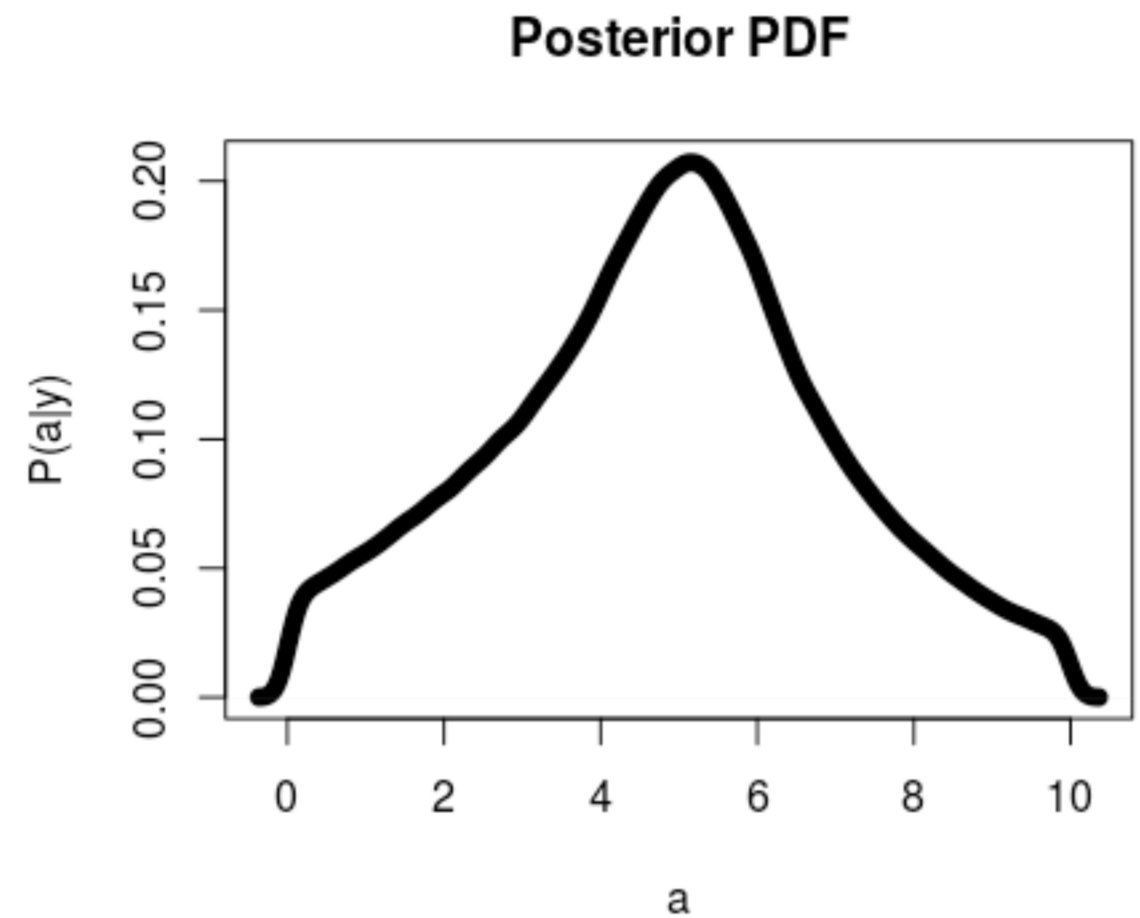


Accept the proposed a_1 with probability u based on a Bernouilli trial:

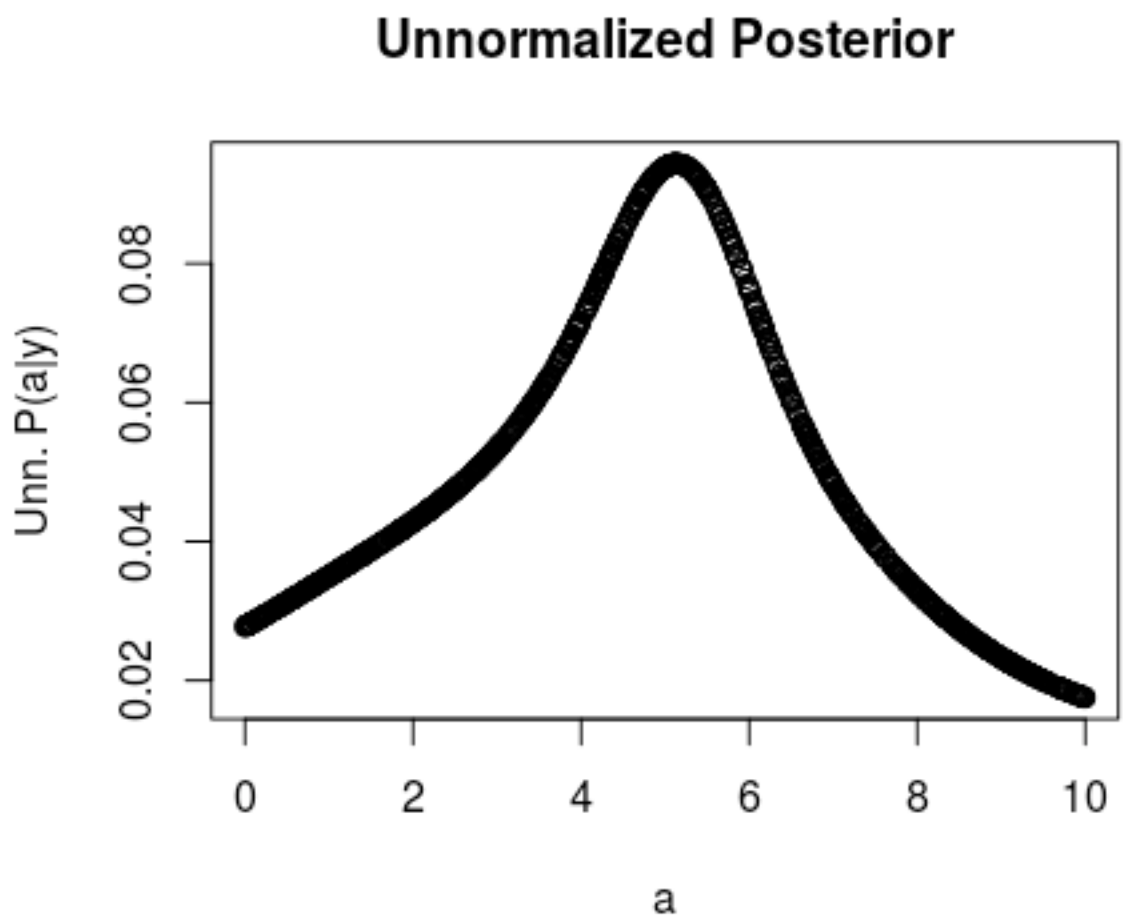
```
if
  runif(1,0,1) < u
  accept
else
  reject
```

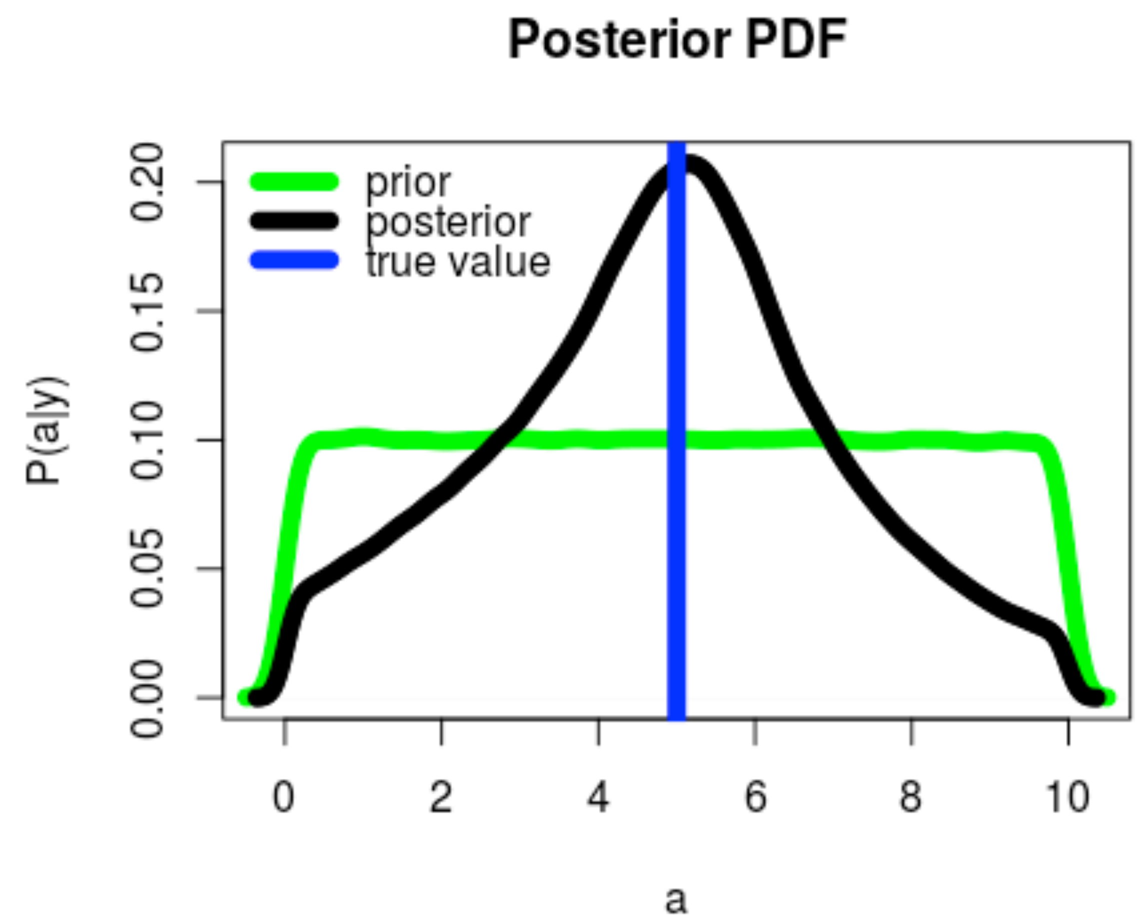
Rejection sampling

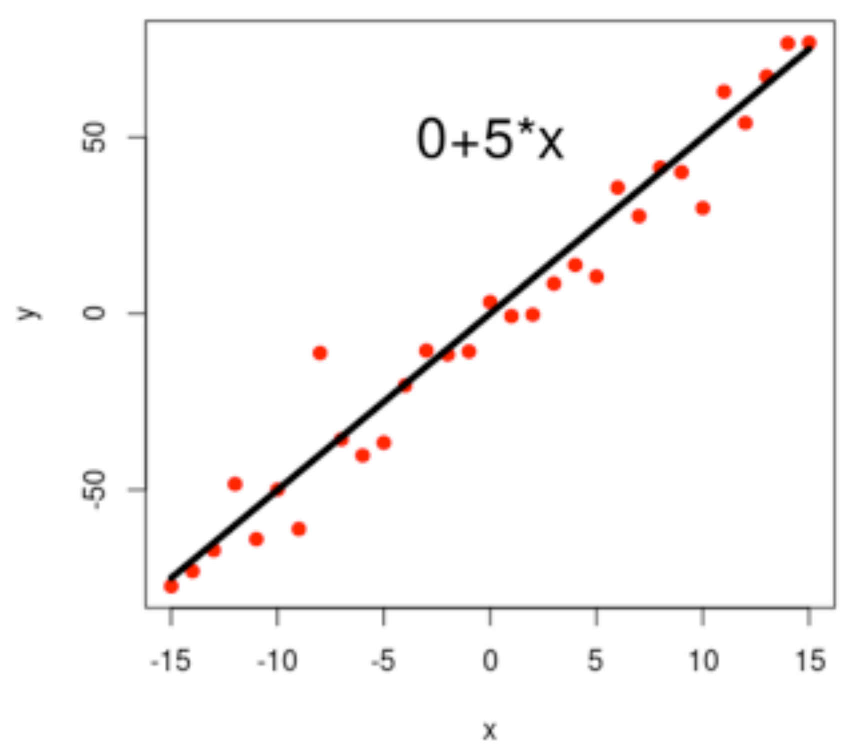




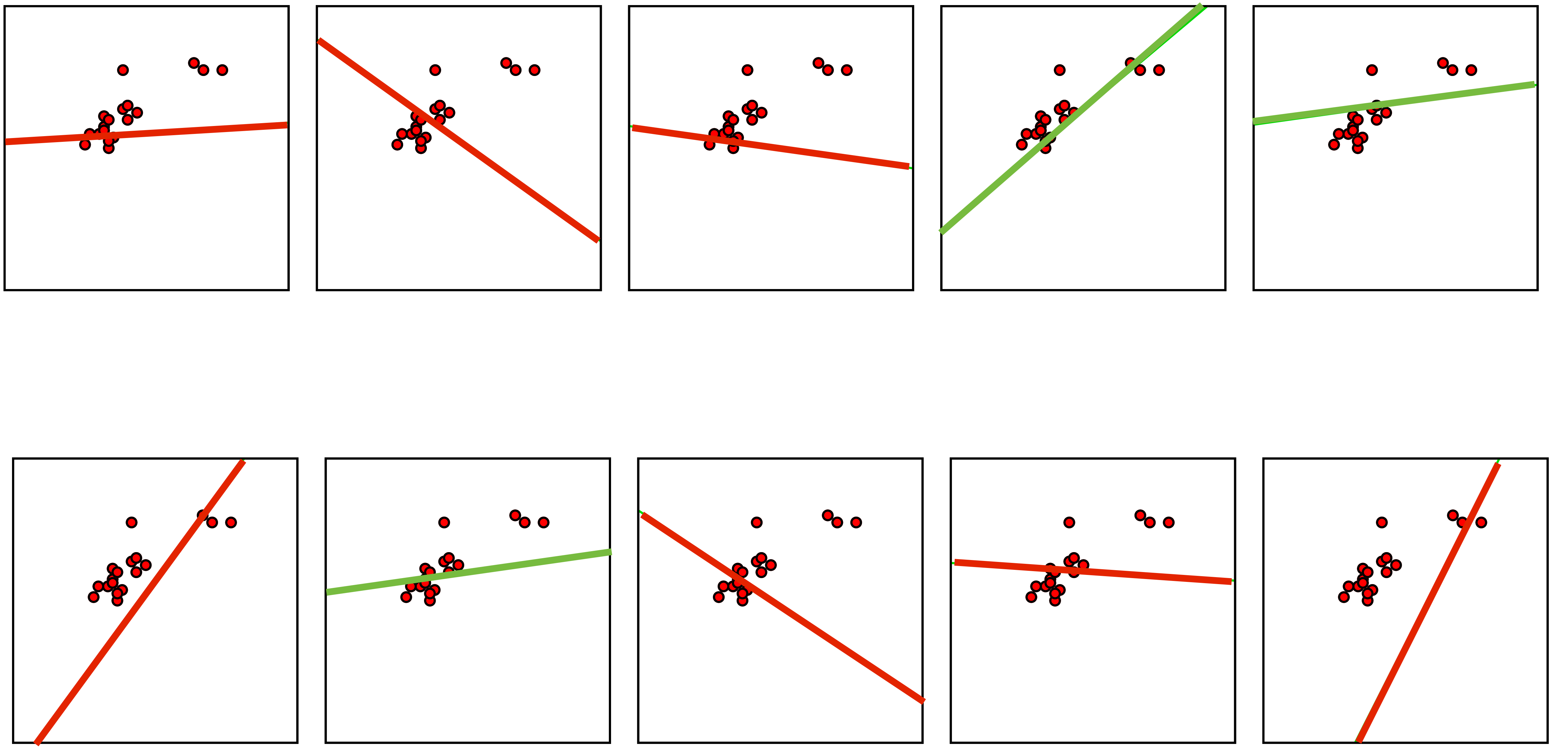
Rejection sampling







Rejection sampling

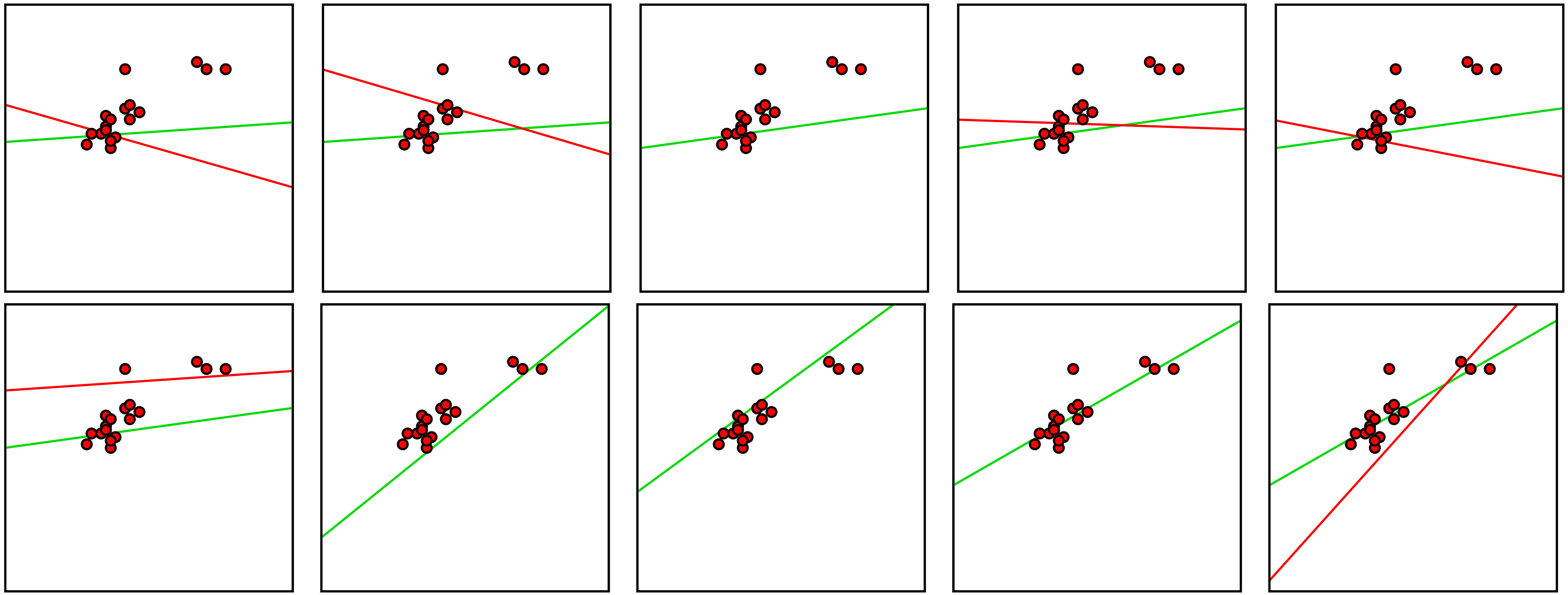


Rejection sampling

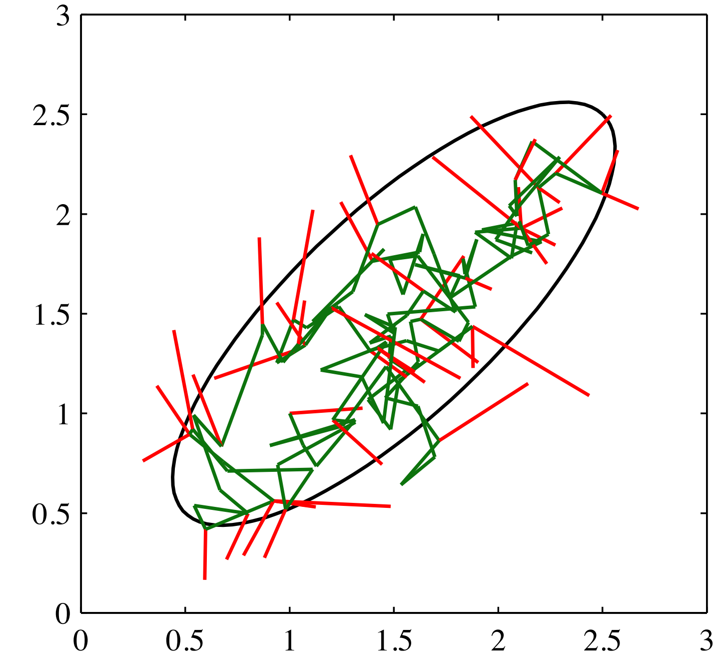
Pros:
- Parallelizable
- Easy to implement

Cons:
- Suffers from curse of dimensionality

Markov Chain Monte Carlo



- Perturb parameters
- Accept if new params are supported more (*also accept sometimes even if not)
- Otherwise keep old parameters



Markov Chain Monte Carlo

1) Start from some initial parameter value
2) Evaluate the unnormalized posterior
3) Propose a new parameter value
4) Evaluate the new unnormalized posterior
5) Decide whether or not to accept the new value
6) Repeat 3-5

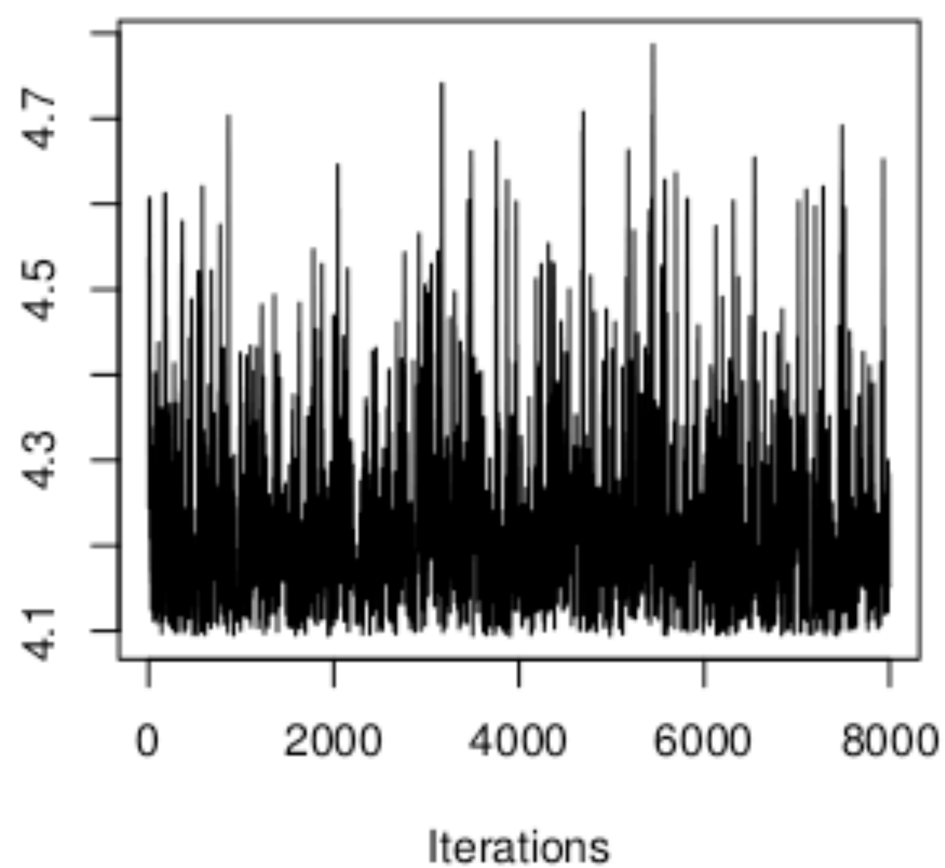
Markov Chain Monte Carlo

1) Start from some initial parameter value
2) Evaluate the unnormalized posterior
3) Propose a new parameter value → **dependent on step 1/previous step**
4) Evaluate the new unnormalized posterior
5) Decide whether or not to accept the new value
6) Repeat 3-5

- Advantages
  - Multi-dimensional
  - Can be applied to
    - Whole joint PDF
    - Each dimension iteratively
    - Groups of parameters
  - Simple
  - Robust
- Disadvantages
  - Sequential samples not independent
  - Computationally intensive
  - Discard “Burn – in” period before convergence
  - Assessing convergence

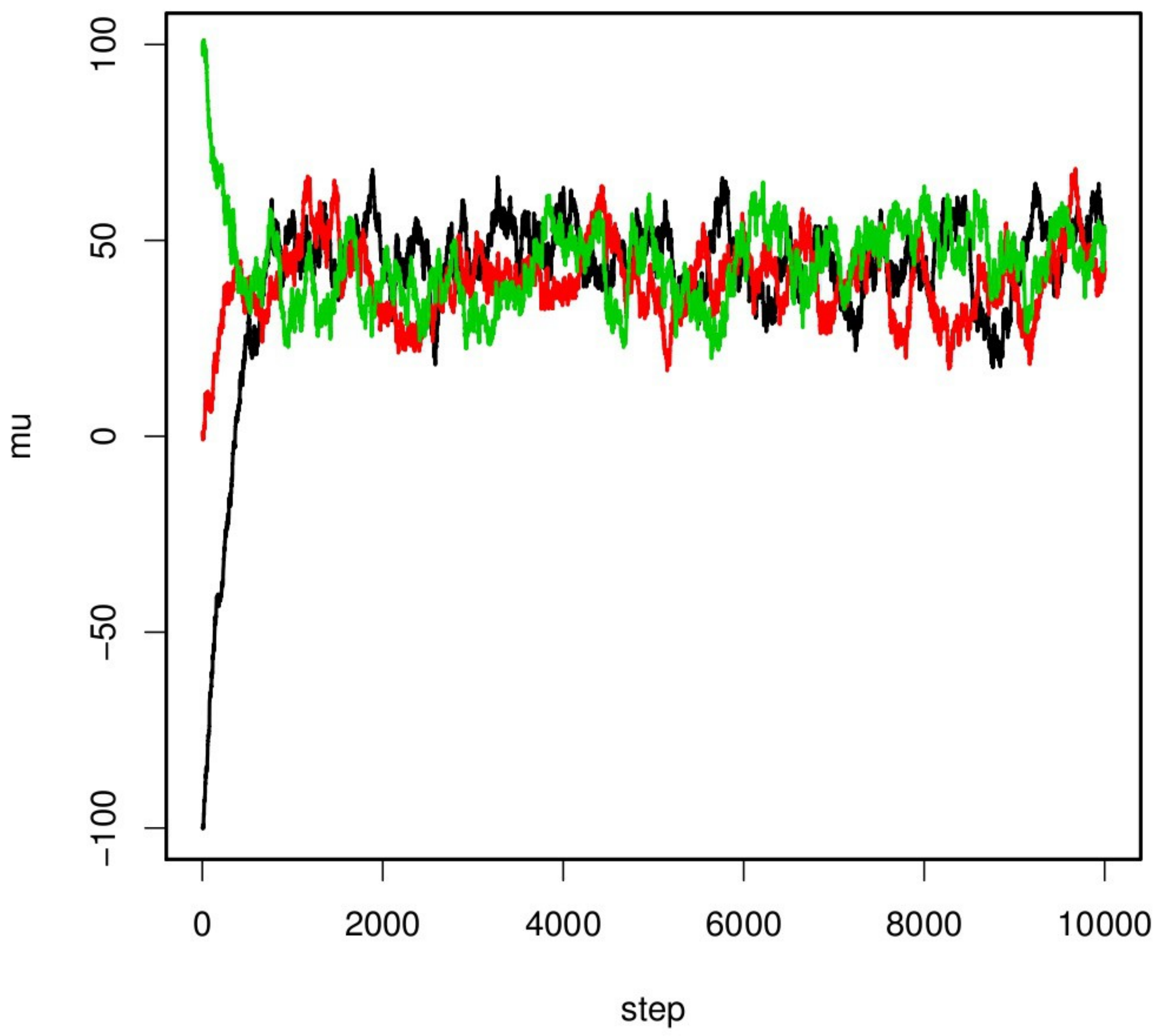
How to assess convergence?

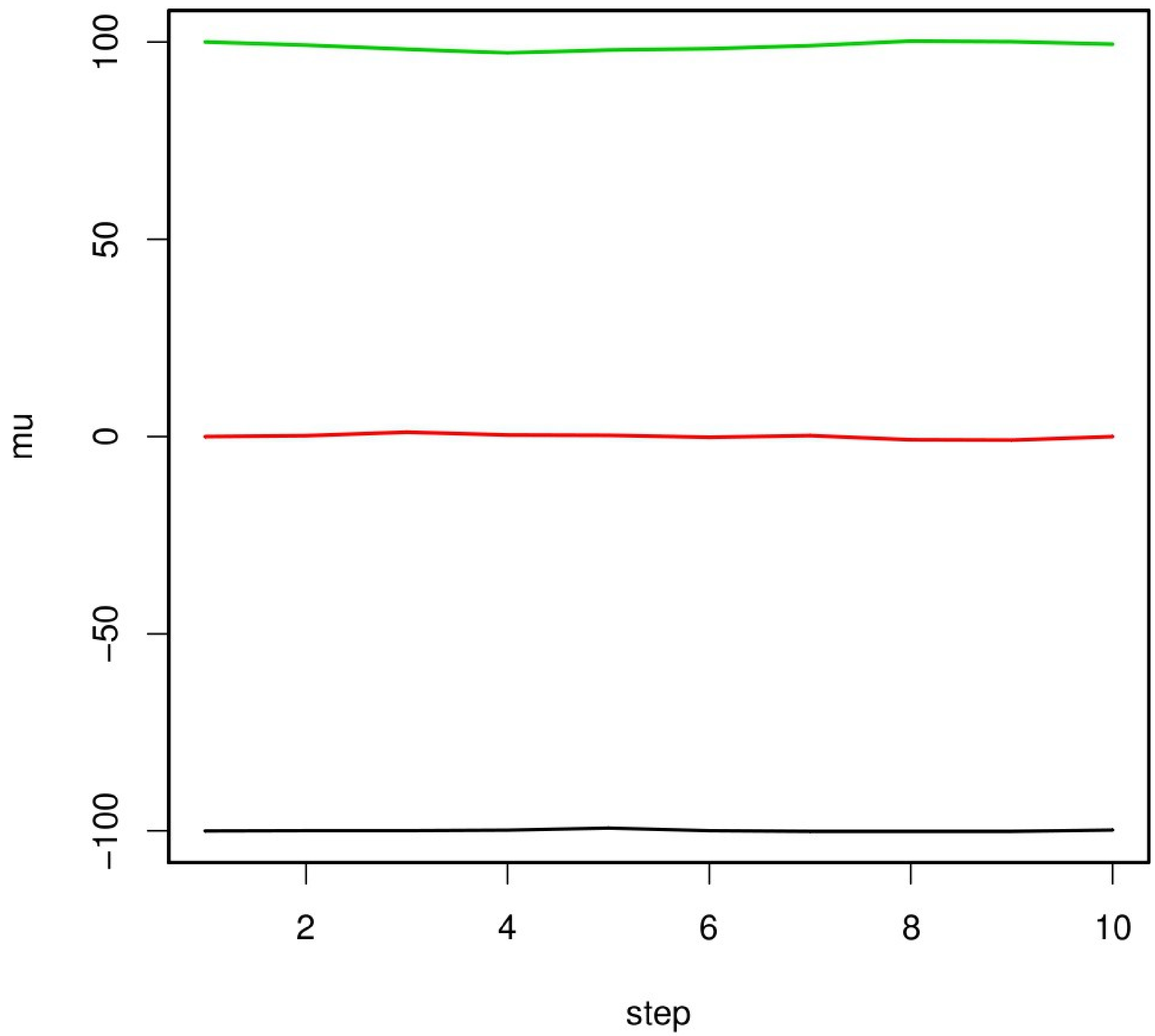
- Visual inspection



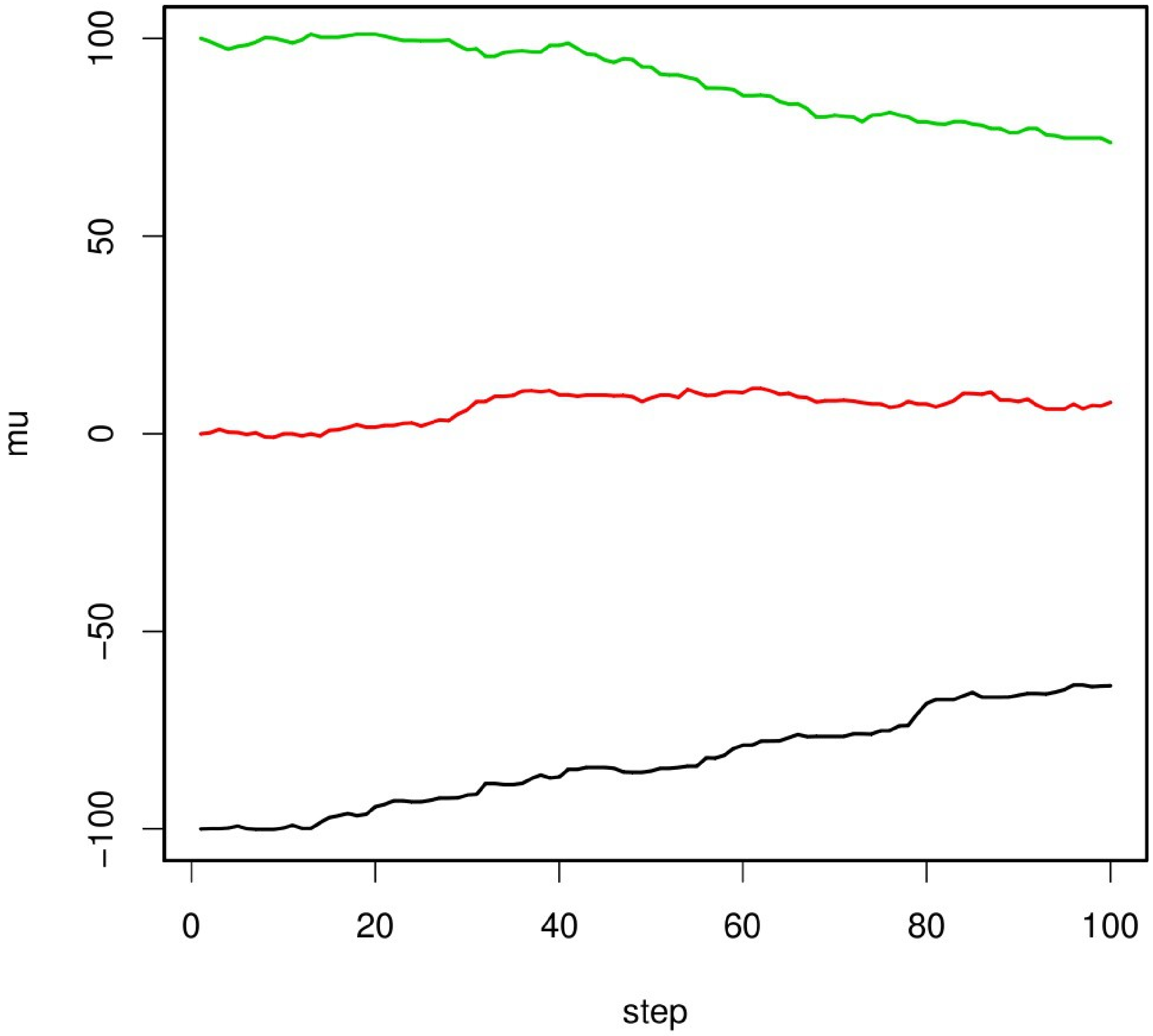
How to assess convergence?

- Visual inspection
- Multiple chains

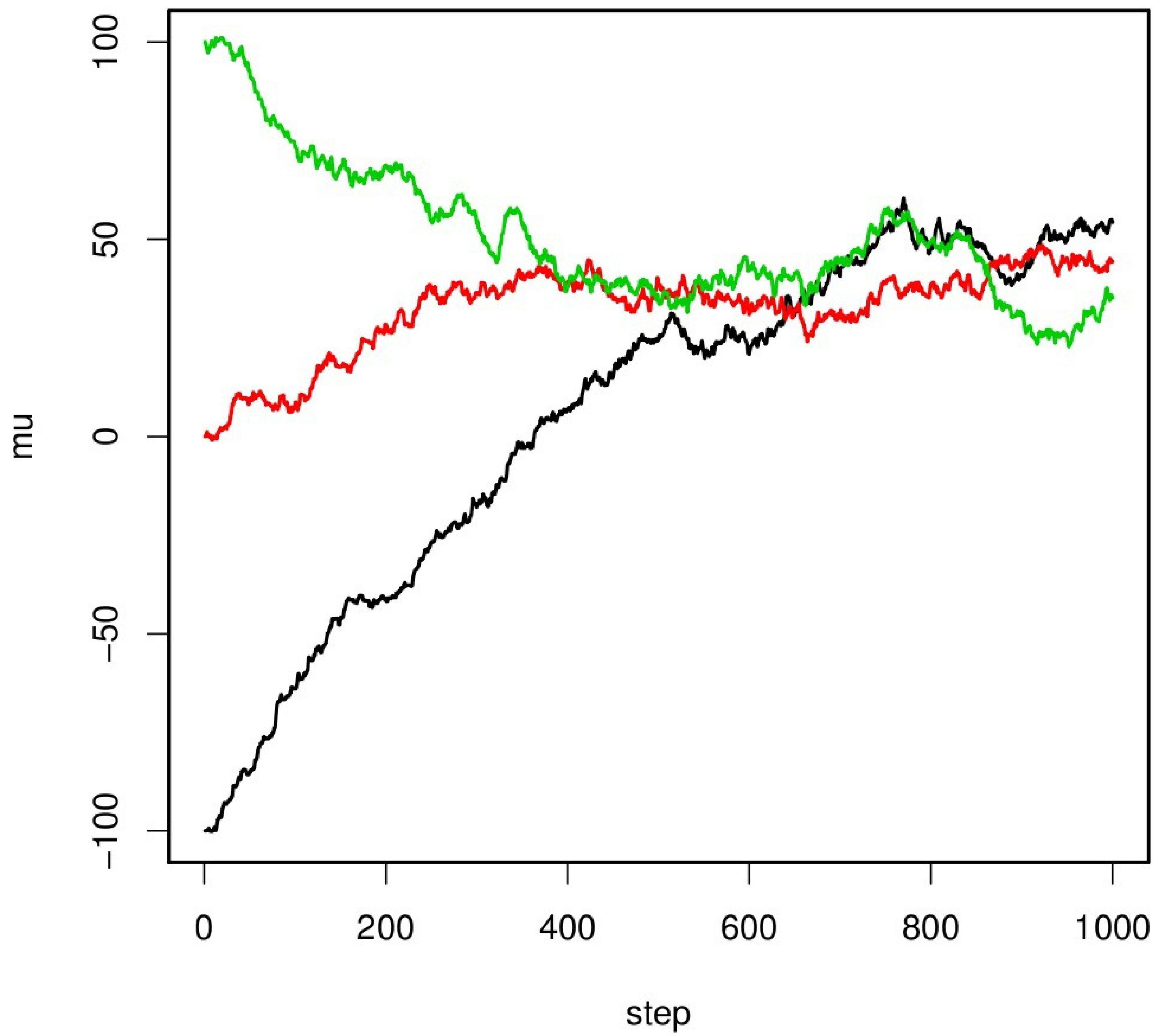




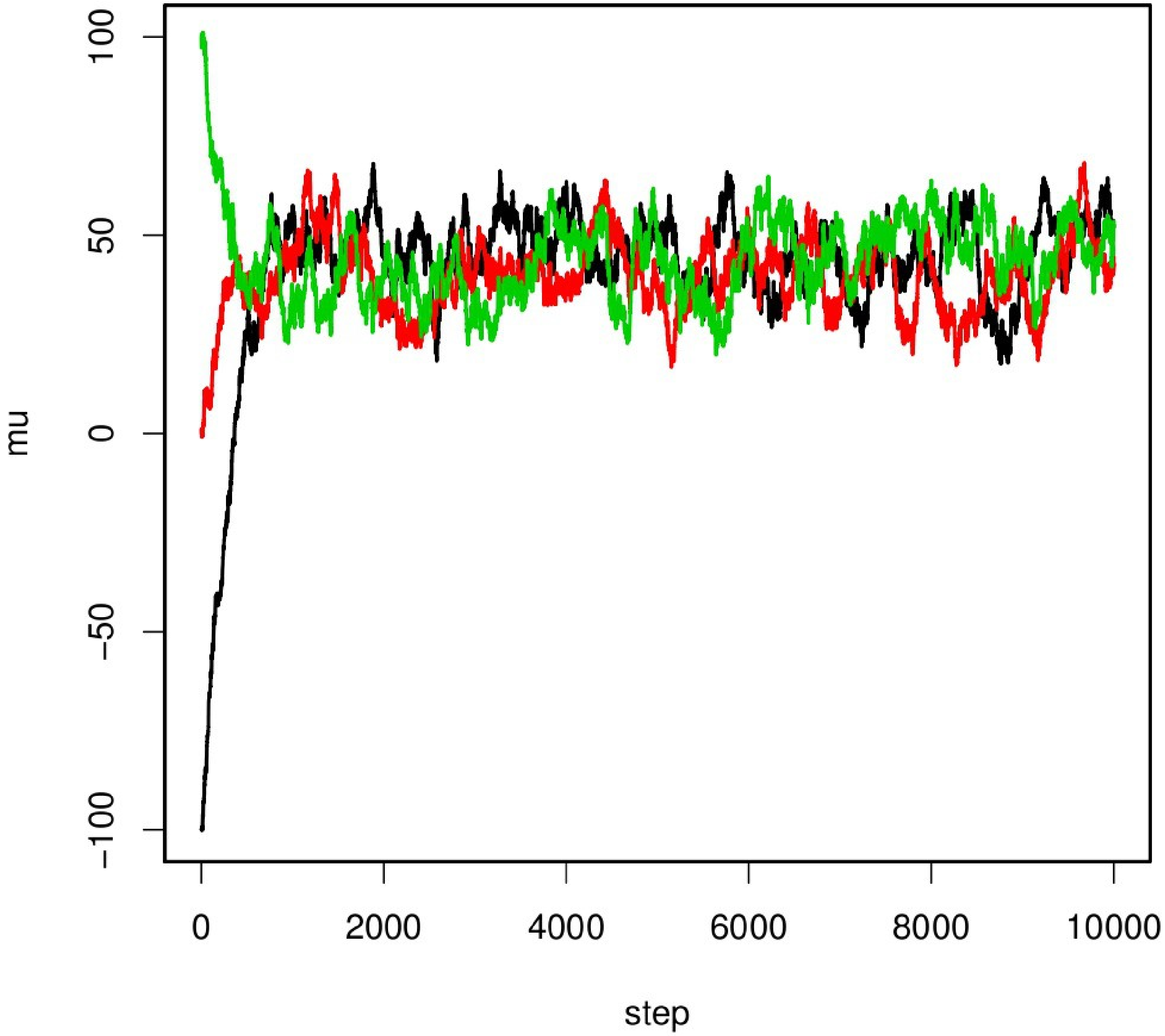
100
50
0
−50
−100
mu
2
4
6
8
10
step



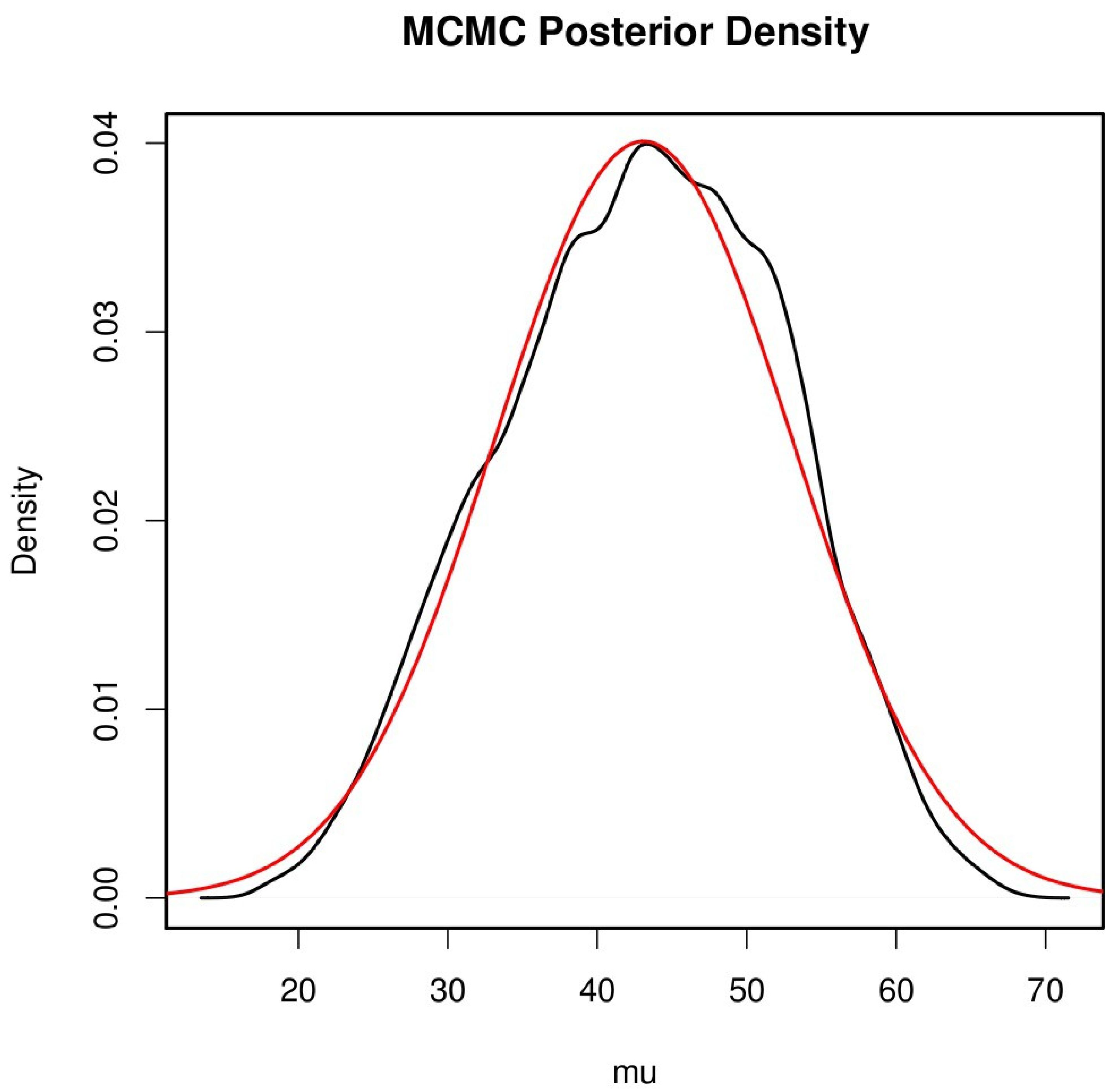
mu
100
50
0
−50
−100
0
20
40
60
80
100
step



100
50
0
−50
−100
mu
0
200
400
600
800
1000
step



100
50
0
−50
−100
mu
0
2000
4000
6000
8000
10000
step


MCMC Posterior Density
Density
0.00
0.01
0.02
0.03
0.04
20
30
40
50
60
70
mu

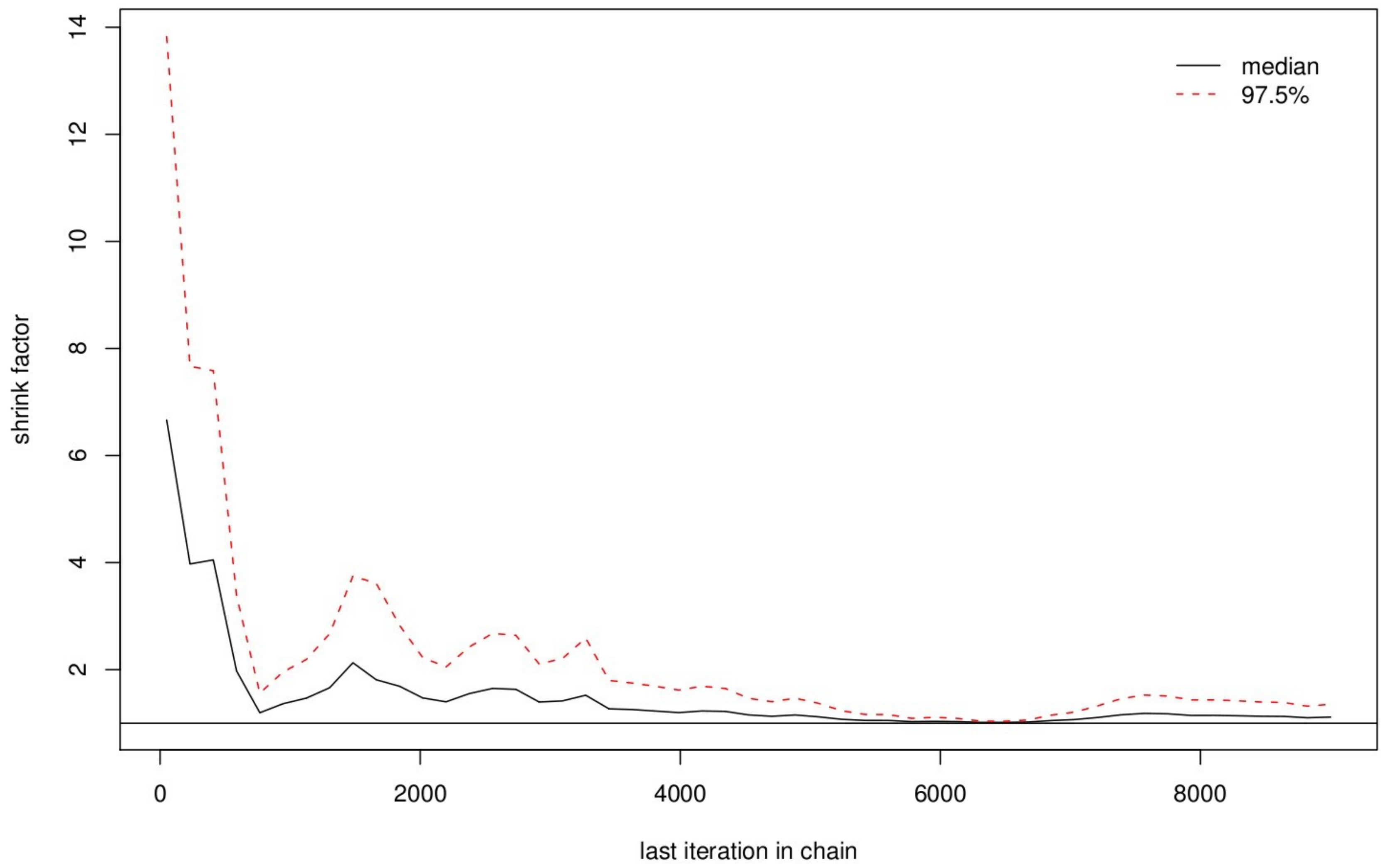
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats

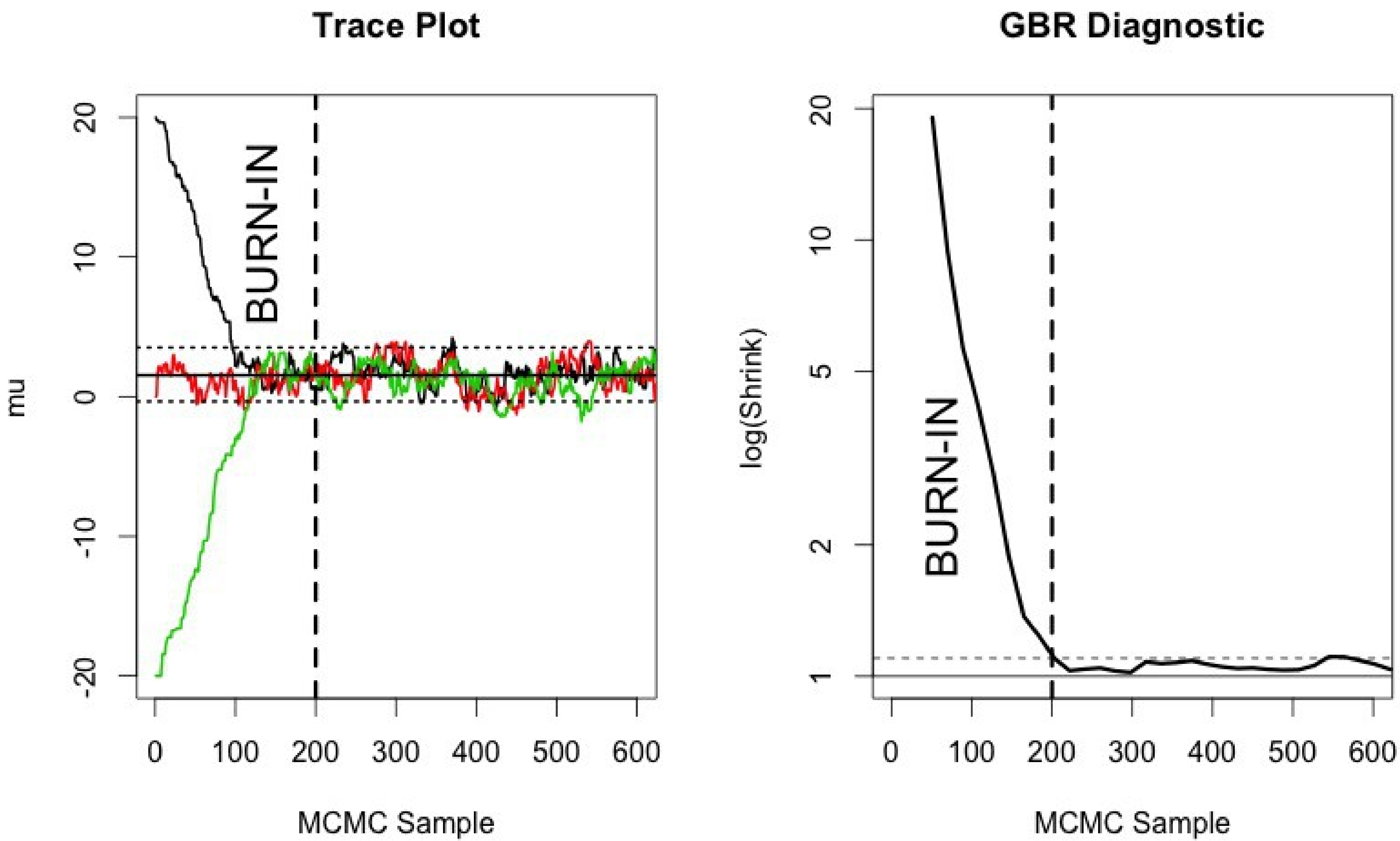
Convergence Statistics

$$\hat{R} = \frac{B}{W}$$

- Brooks Gelman Rubin
  - Within vs among chain variance
  - Should converge to 1



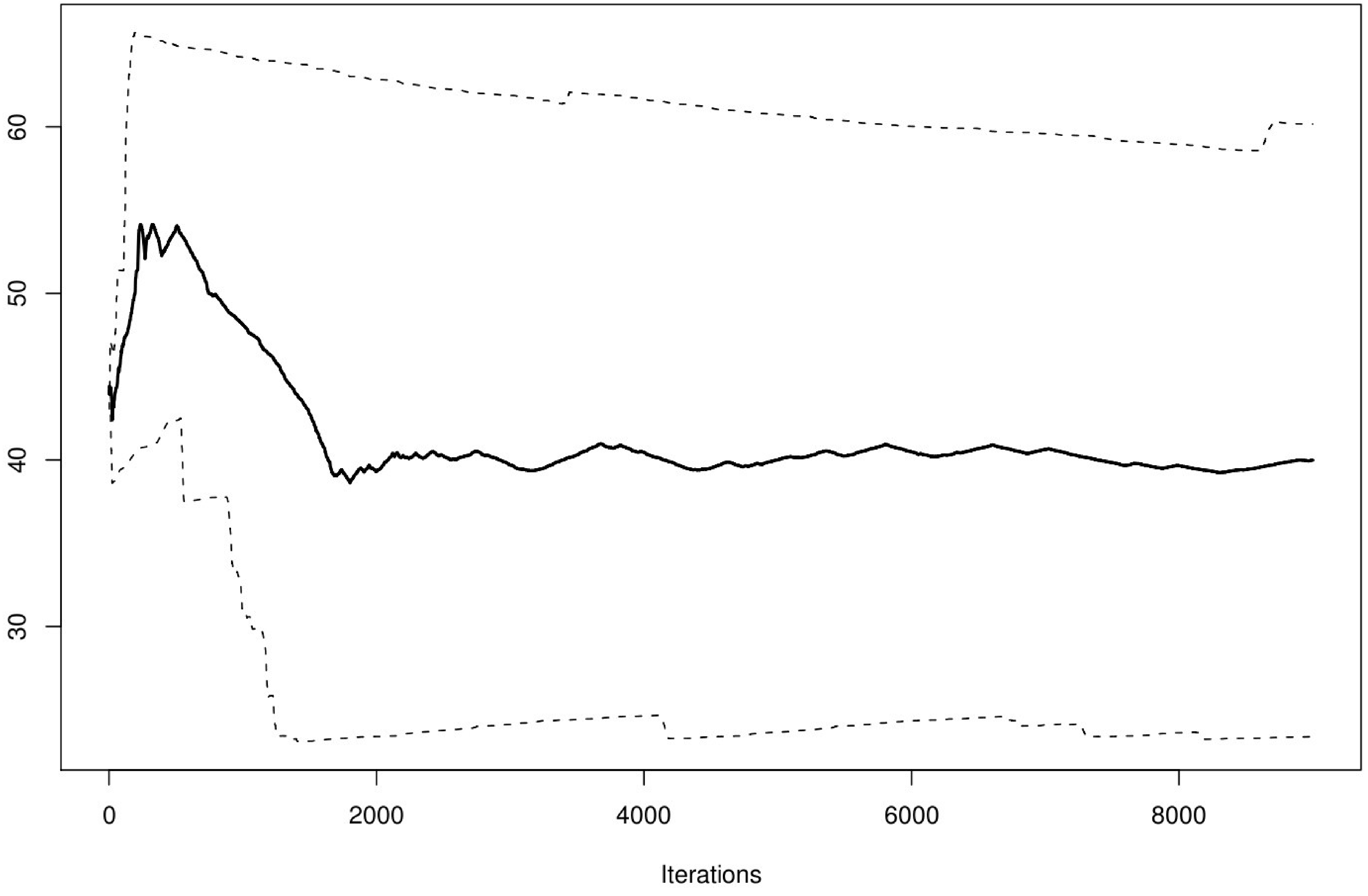
Convergence Statistics



How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles

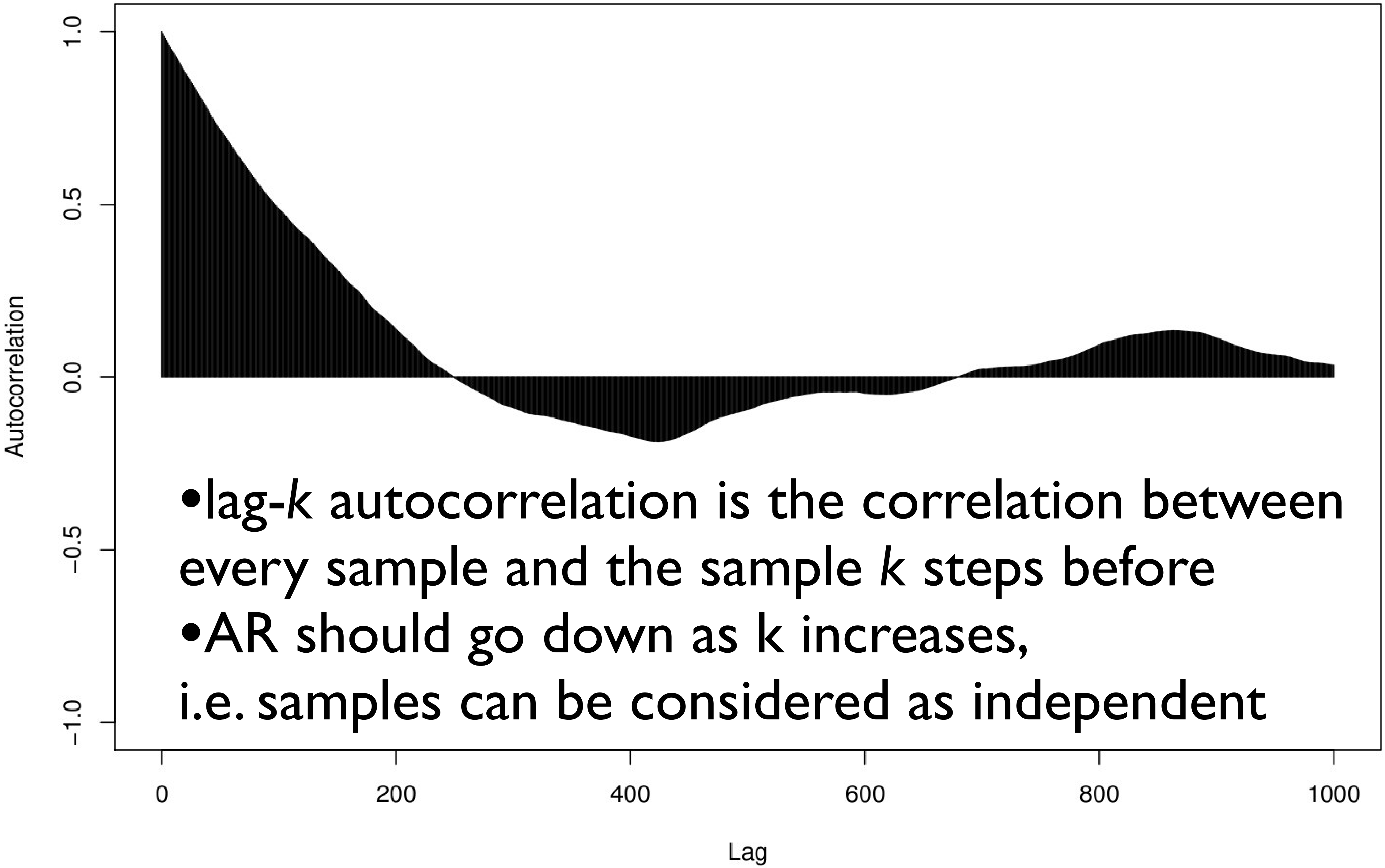
Quantiles



How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation

Autocorrelation



- lag-*k* autocorrelation is the correlation between every sample and the sample *k* steps before
- AR should go down as k increases,
i.e. samples can be considered as independent

How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation
- Summary statistics

Summary Statistics

Analytical:

Mean	SD
43.09901	9.95037

MCMC:

Mean	SD	Naive SE	Time-series
43.05504	9.28108	0.05648	0.74503

Quantiles:

2.5%	25%	50%	75%	97.5%
24.98	36.46	43.39	49.99	60.01

How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation
- Summary statistics
- Effective sample size
- Acceptance rate