MCMC

Numerical methods for Bayes

Numerical Methods for Bayes

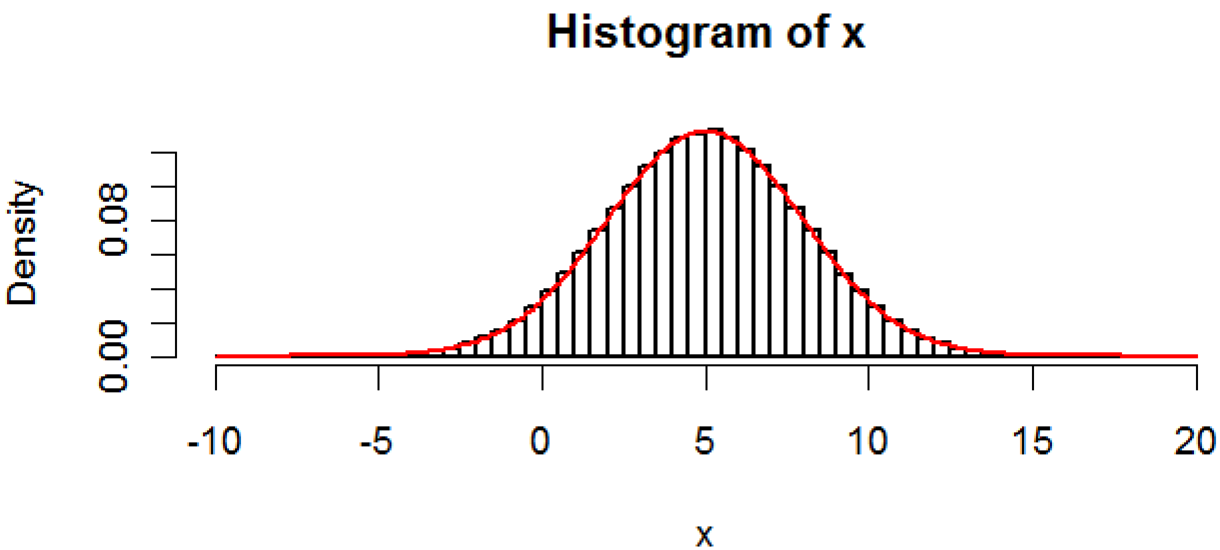
$$P(\theta|y)=\frac{P(y|\theta)P(\theta)}{\int_{-\infty}^{\infty} P(y|\theta)P(\theta)\,d\theta}$$

- <u>Not</u> just optimization
- Need to integrate denominator
  - Numerical Integration
- Would also like to know the mean, median, mode, variance, quantiles, confidence intervals, etc.

Idea:

Random samples from the posterior

- Approximate PDF with the histogram
- Performs *Monte Carlo Integration*
- Allows all quantities of interest to be calculated from the sample (mean, quantiles, var, etc)



	TRUE	Sample
mean	5.000	5.000
median	5.000	5.004
var	9.000	9.006
Lower CI	-0.880	-0.881
Upper CI	10.880	10.872

Outline

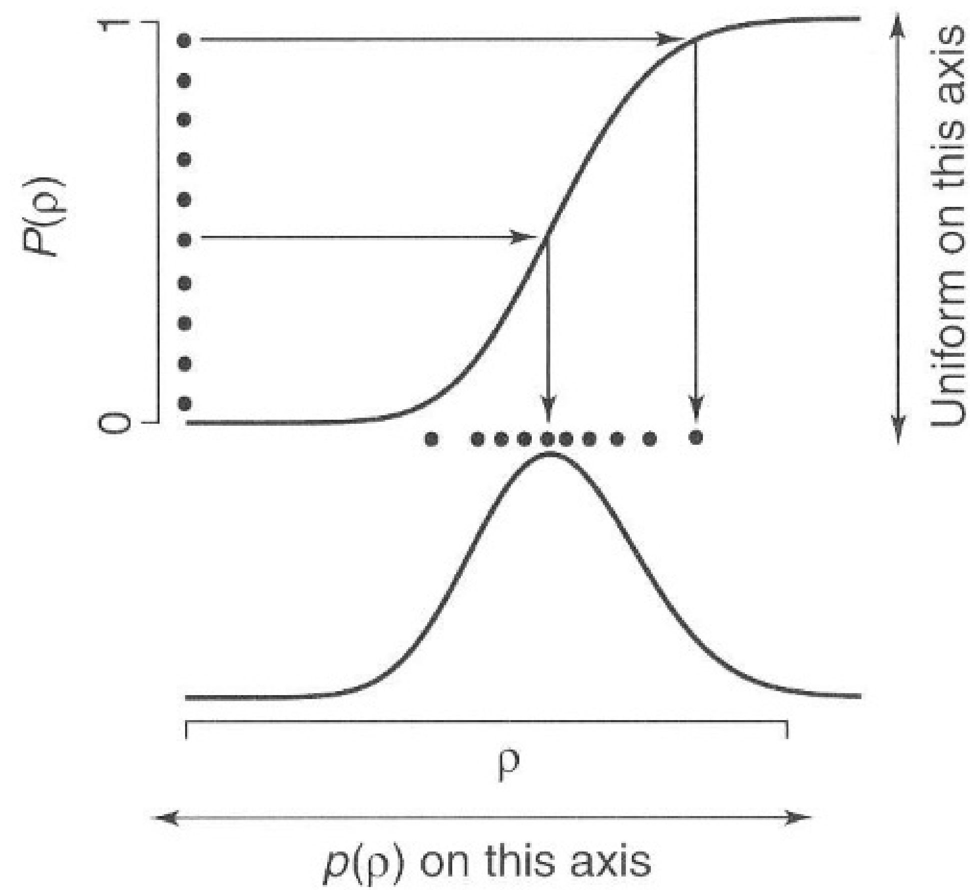
- **Different numerical techniques for sampling from the posterior**
  - **Inverse Distribution Sampling**
  - **Rejection Sampling & SMC**
  - **Markov Chain-Monte Carlo (MCMC)**
    - Metropolis
    - Metropolis-Hastings
    - Gibbs sampling
- Sampling conditionals vs full model
- Flexibility to specify complex models

How do we generate a random number from a PDF?

- Exist for most standard distributions
- Posteriors often non-standard
- Indirect Methods
  - First sample from a different distribution
  - Rejection sampling, Metropolis, M-H
- Direct Methods
  - Inverse CDF
  - Univariate sampling of multivariate or conditional

Inverse CDF sampling

1) Sample from a uniform distribution
2) Transform sample using inverse of CDF, $F^{-1}(x)$



Example: Exponential

- The exponential CDF is: $F(x) = 1 - e^{-\lambda x}$
- We solve for F^{-1} as

$$p = 1 - e^{-\lambda x}$$

$$1 - p = e^{-\lambda x}$$

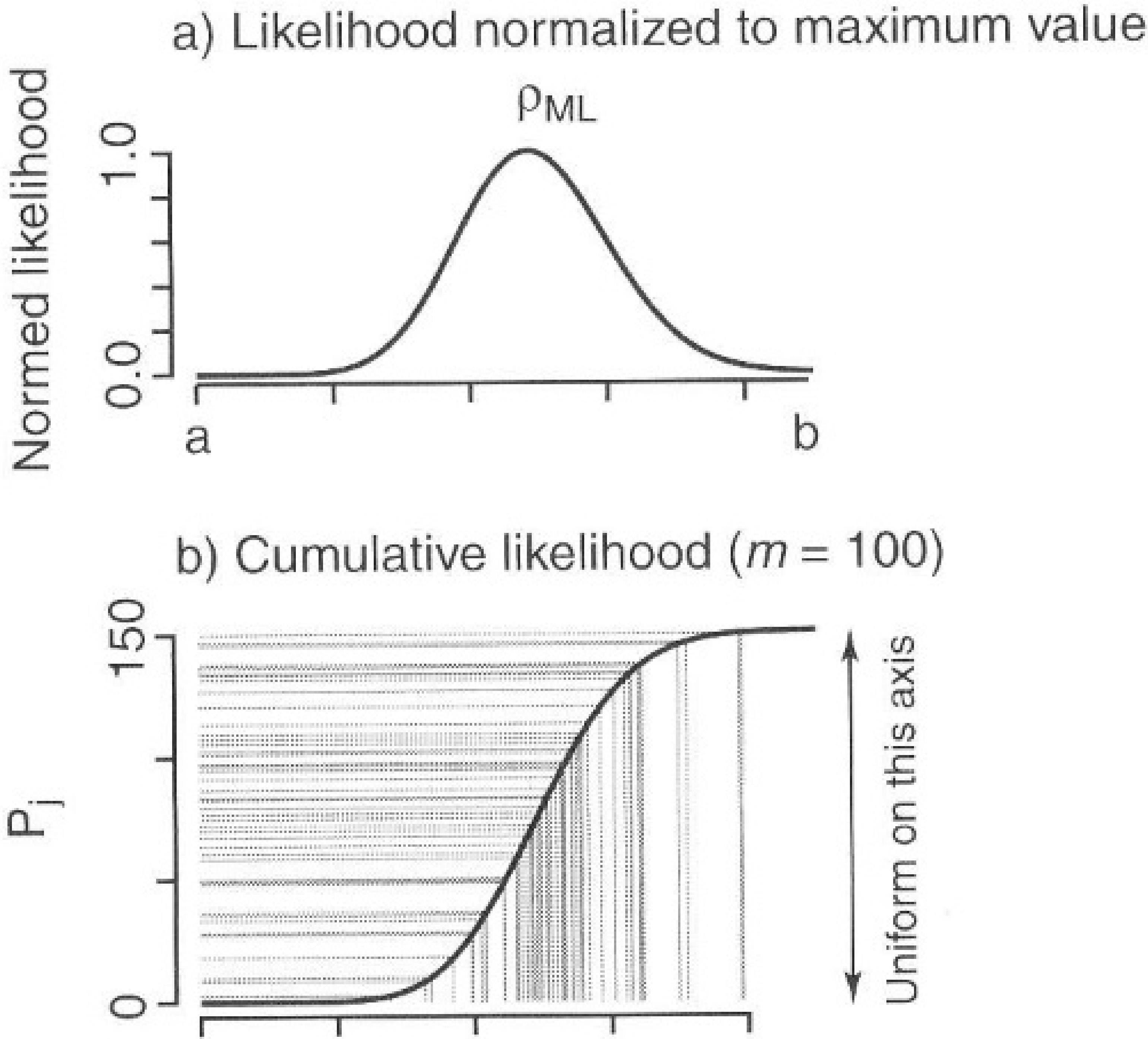
$$\ln(1-p) = -\lambda x$$

$$x = F^{-1}(p) = -\frac{\ln(1-p)}{\lambda}$$

- Draw p ~ Unif(0,1), calculate x

Approximate inverse sampling

- Exact inverse sampling requires CDF & ability to solve for inverse
- Approximation
  - Solve for $f(x)$ across a discrete sequence of x
  - Determine cumulative sum to approx $F(x)$
  - Draw $Z \sim \text{unif}(0,\max)$
  - Find the value of x for which $Z == \text{cumsum}(f(x))$
- Approximation performs integration as a Riemann sum



a) Likelihood normalized to maximum value
Normed likelihood
1.0
0.0
ρ_{ML}
a
b
b) Cumulative likelihood (m = 100)
150
0
P_j
Uniform on this axis

Univariate sampling of multivariate or conditional distribution

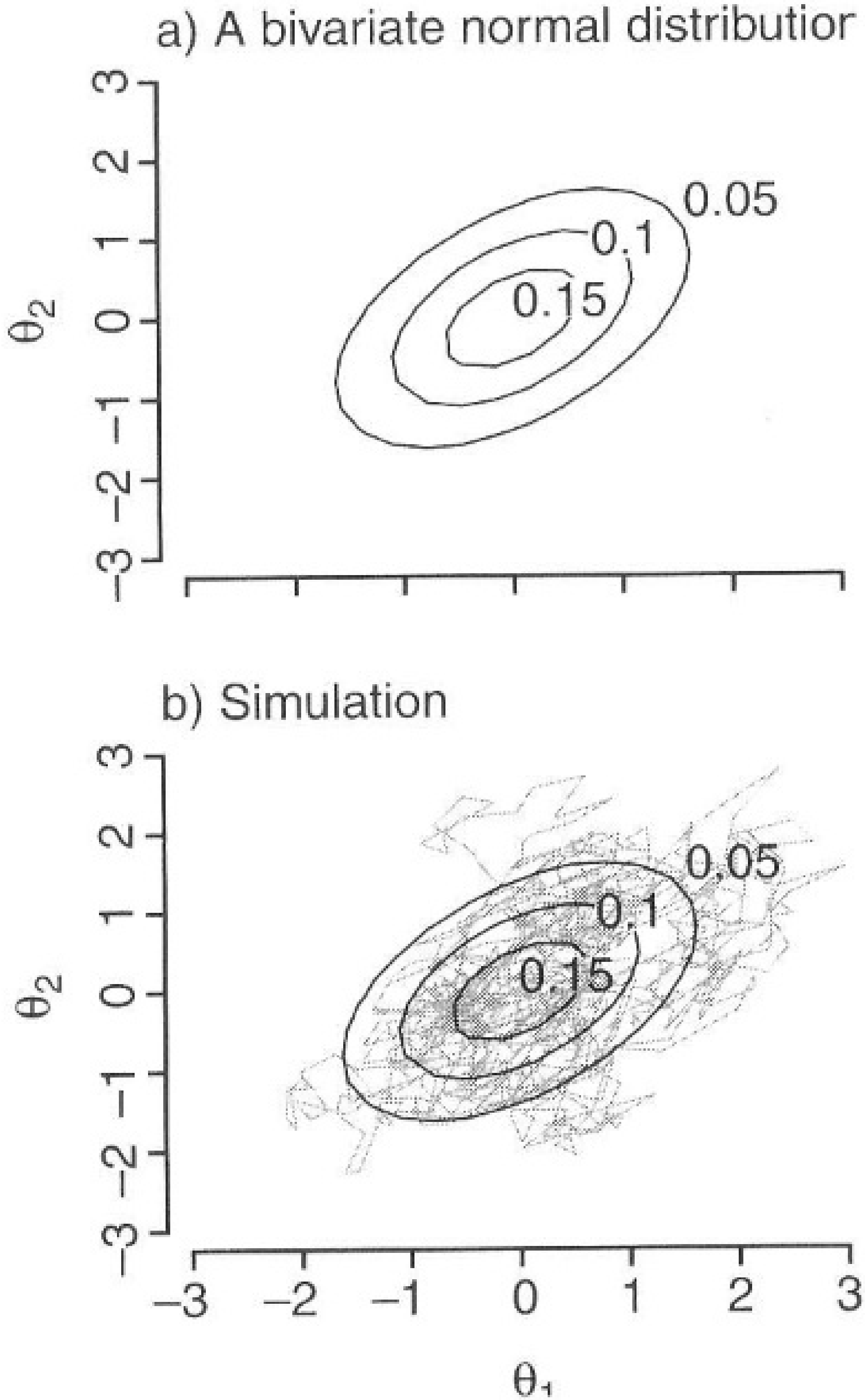
- Multivariate
  - Multivariate normal based on Normal
  - Multinomial based on Binomial
- Conditional
  - Sample from the first distribution
  - Sample from the second conditioned on the first
  - Examples
    - NBin = Pois($y|\lambda$)Gamma(λ |a,b)
    - Students t = Normal(x | μ,σ^2) IG(σ^2|a,b)

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

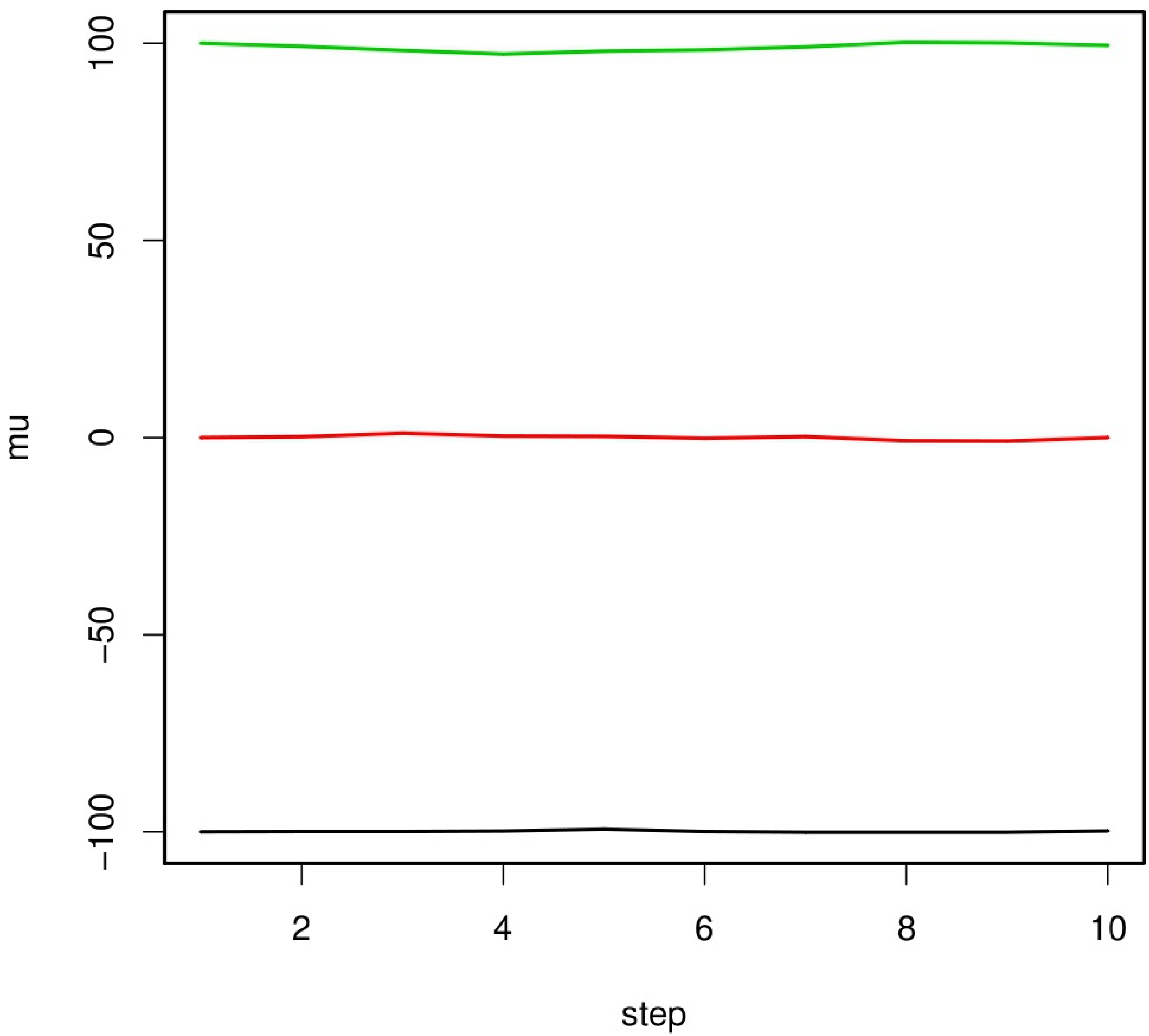
- Looks remarkably similar to optimization
  - Evaluating posterior rather than just likelihood
  - “Repeat” does not have a stopping condition
  - **Criteria for accepting a proposed step**
    - Optimization – diverse variety of options but no “rule”
    - MCMC – stricter criteria for accepting
- Performs random walk through PDF
- Converges “in distribution” rather than to a single point



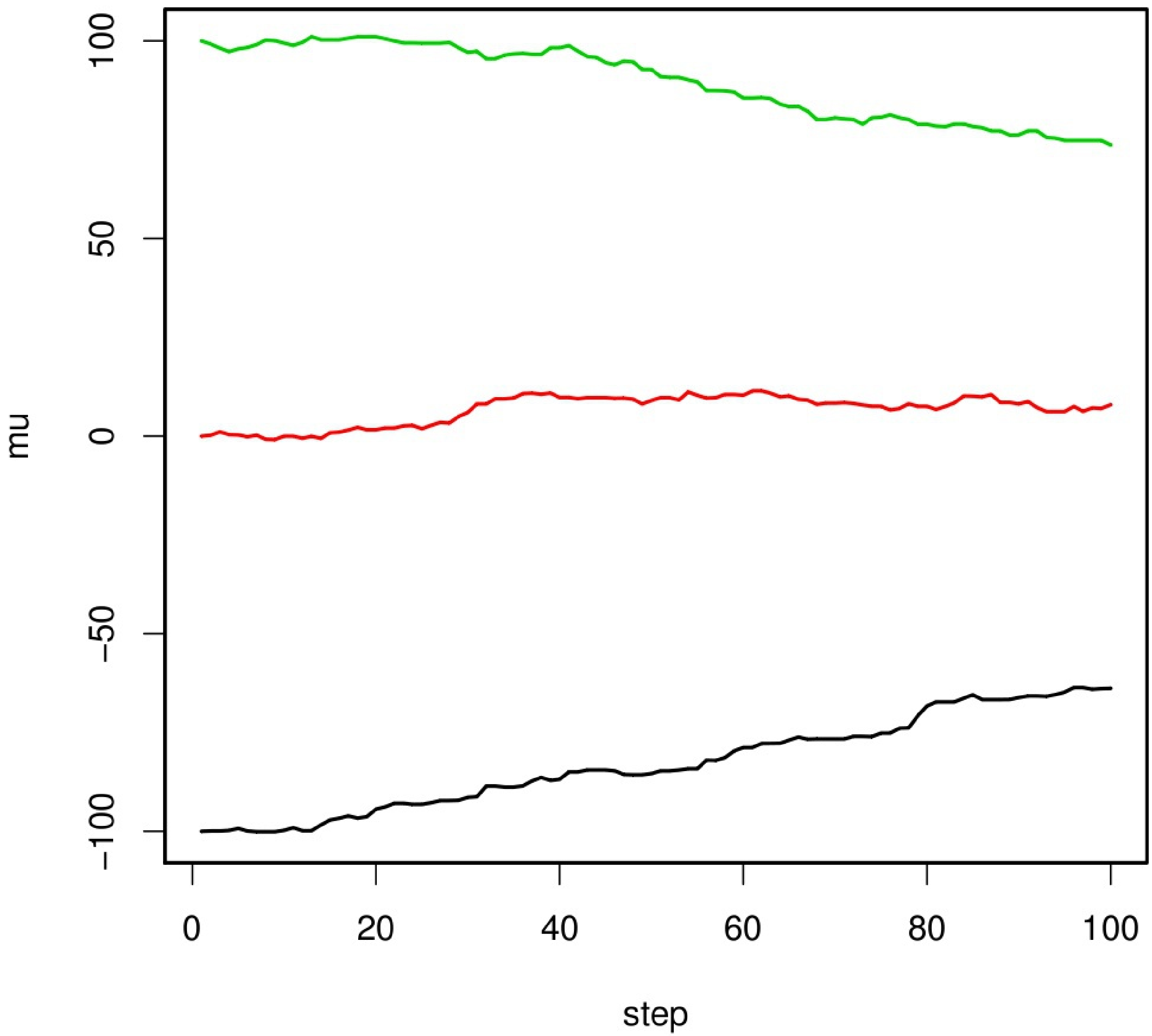
a) A bivariate normal distributior
b) Simulation
0.05
0.1
0.15
θ2
θ1
−3
−2
−1
0
1
2
3

Example

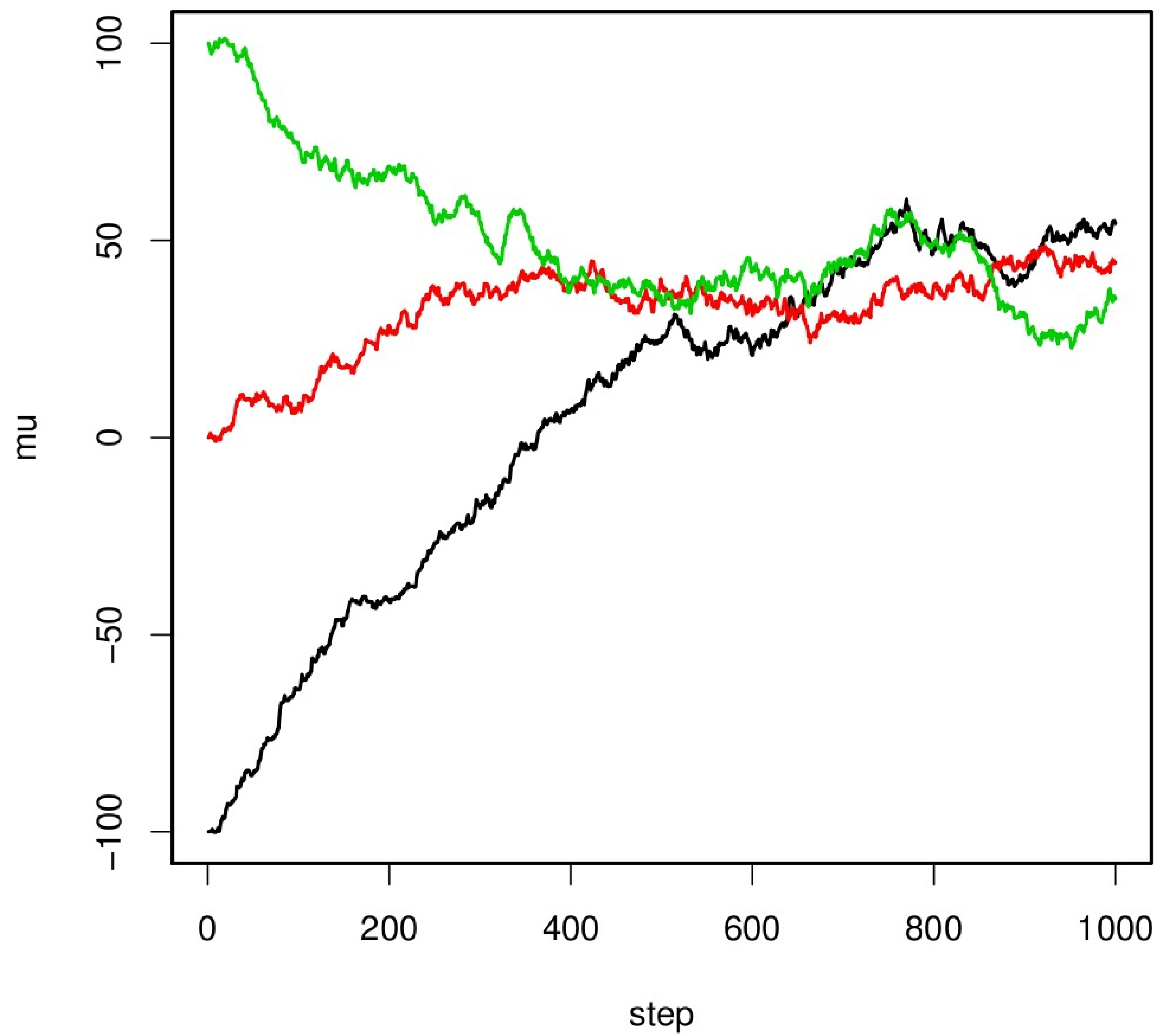
- Normal with known variance, unknown mean
  - Prior: N(53,10000)
  - Data: y = 43
  - Known variance: 100
  - Initial conditions, 3 chains starting at -100, 0, 100



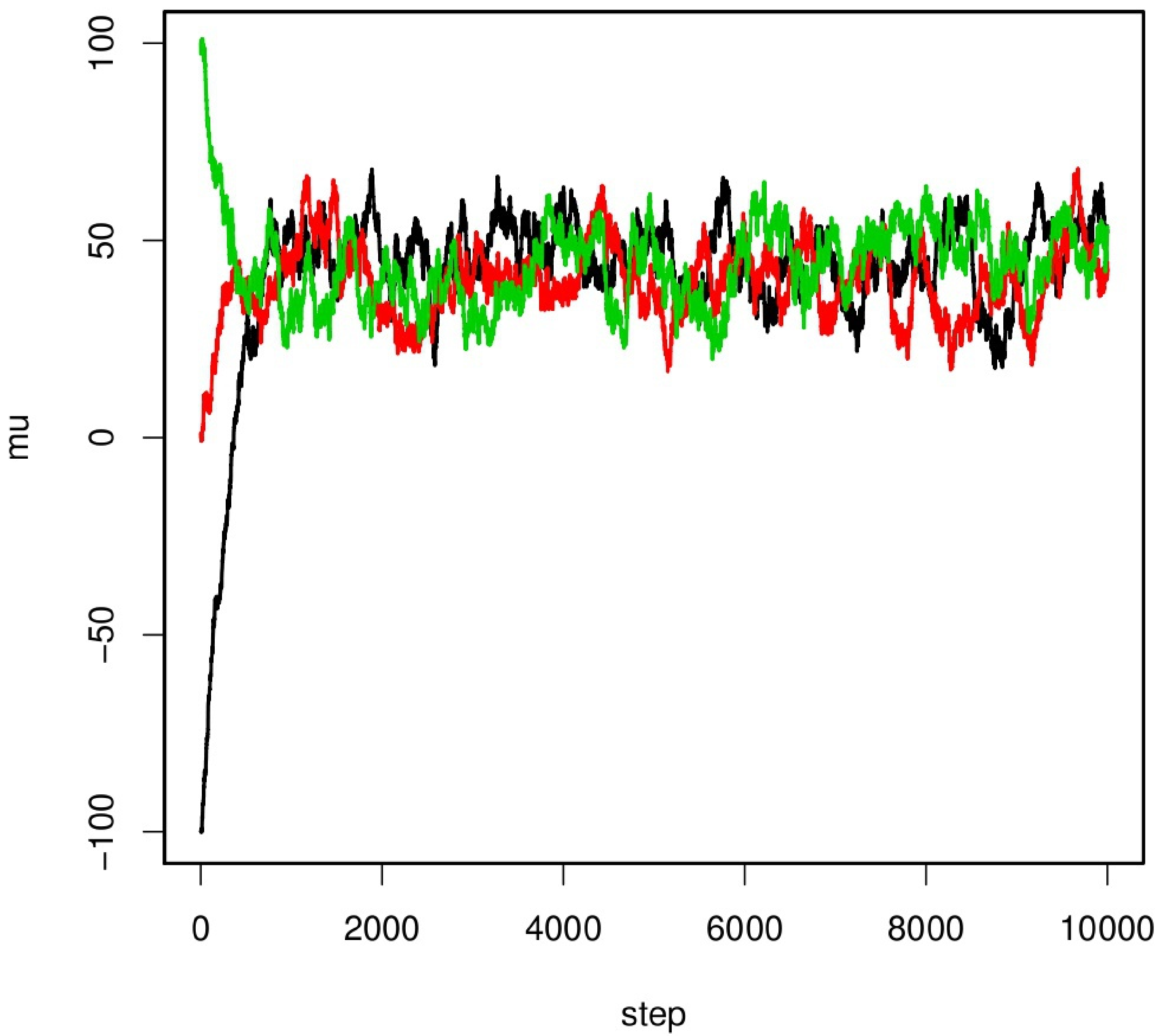
100
50
0
−50
−100
mu
2
4
6
8
10
step



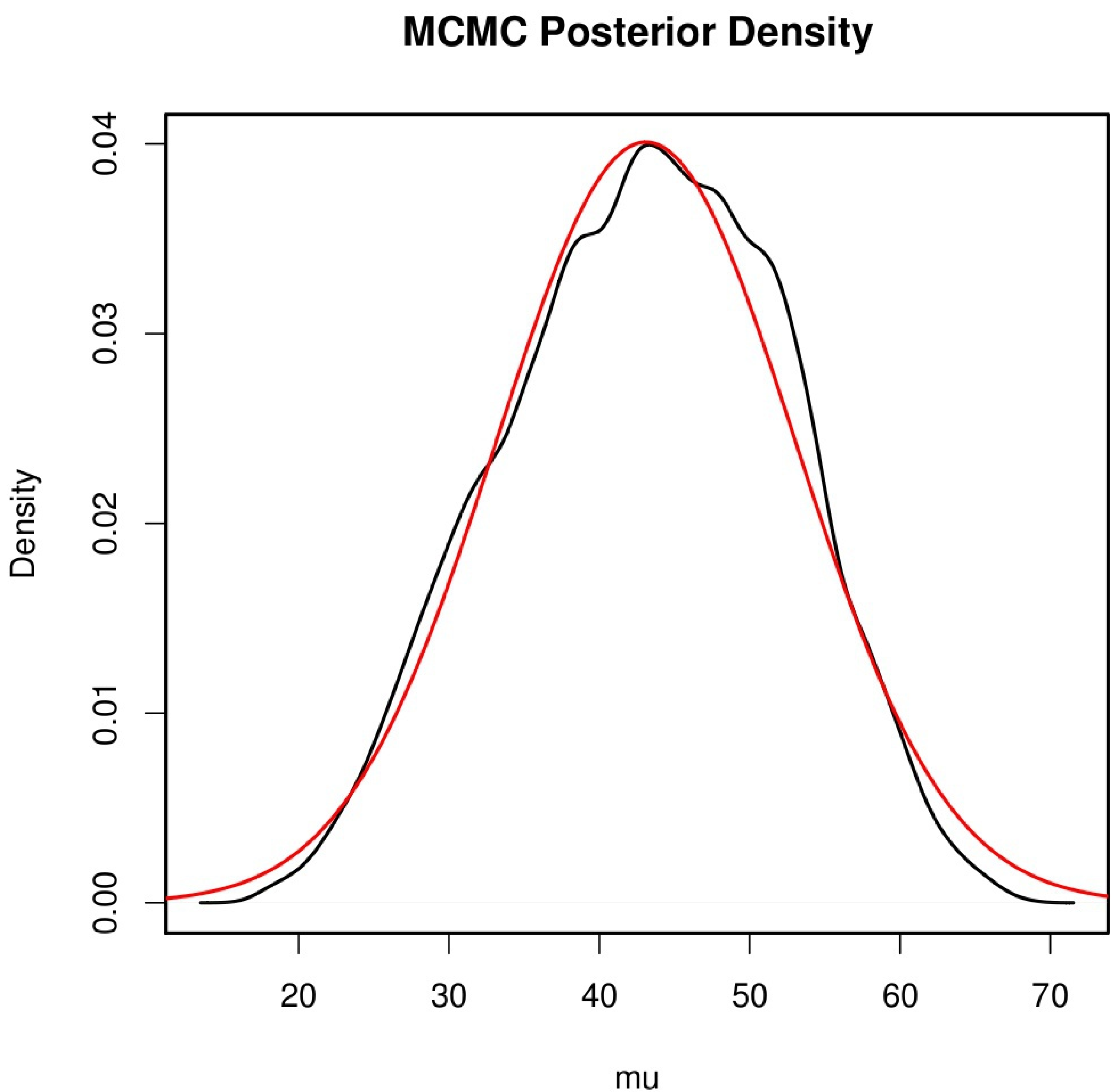
mu
100
50
0
−50
−100
0
20
40
60
80
100
step



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



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

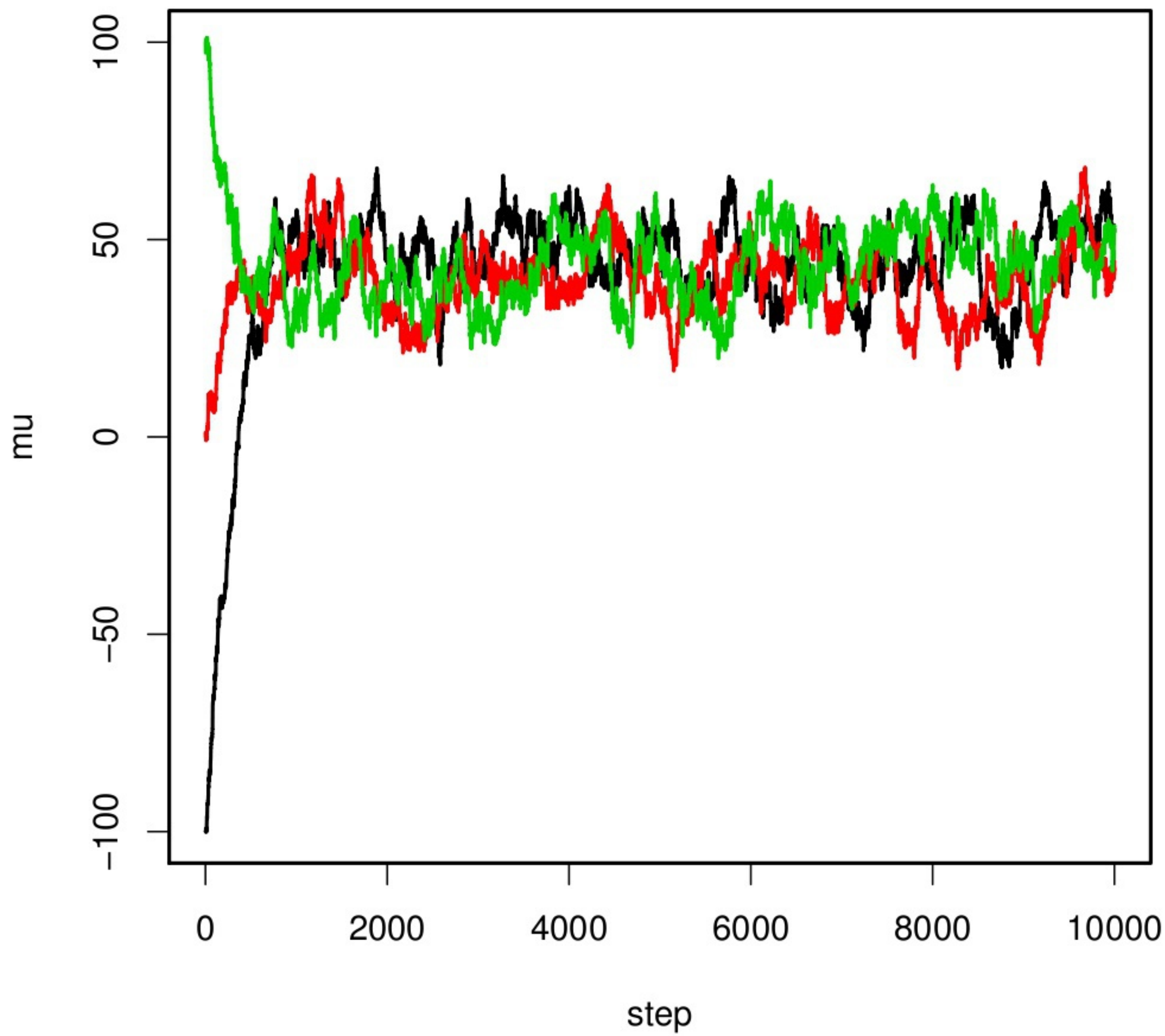

MCMC Posterior Density
Density
0.00
0.01
0.02
0.03
0.04
20
30
40
50
60
70
mu

- 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

Convergence

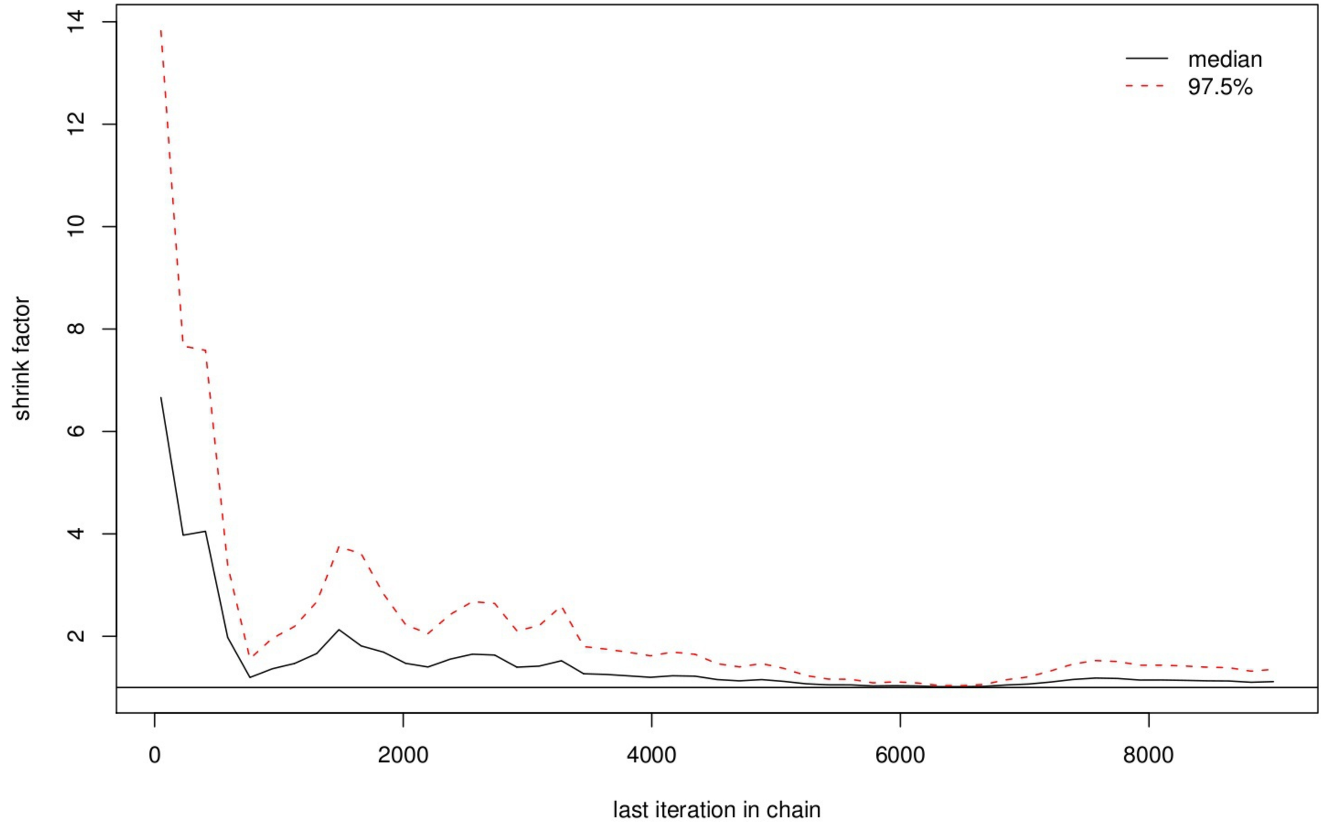
- Generally can not be "proved"
- Why MCMC can be "dangerous," especially in the hands of the untrained
- Assessed by examining MCMC time-series
  - Visual inspection
  - Multiple chains
  - Convergence statistics
  - Acceptance rate
  - Auto-correlation

Visual inspection / multiple chains

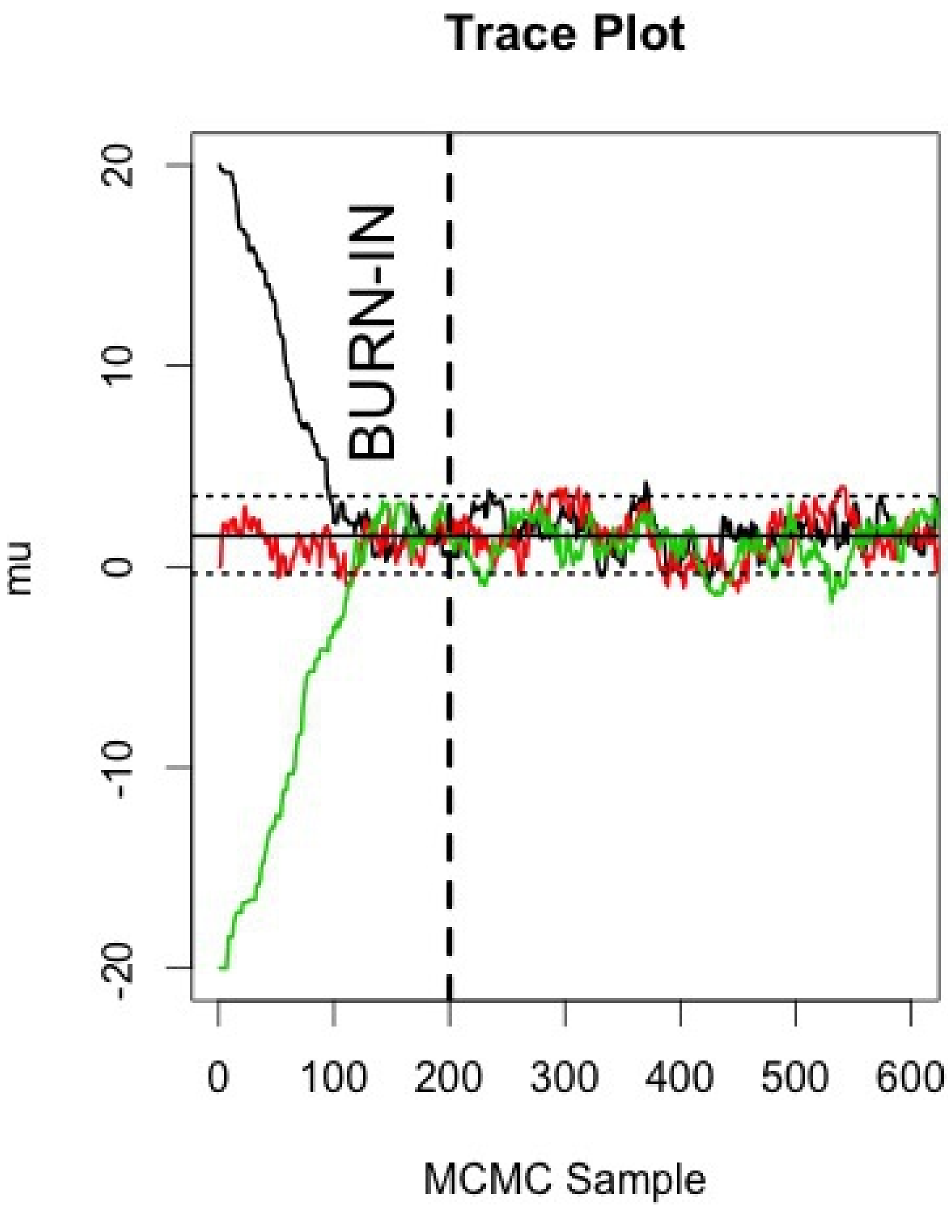


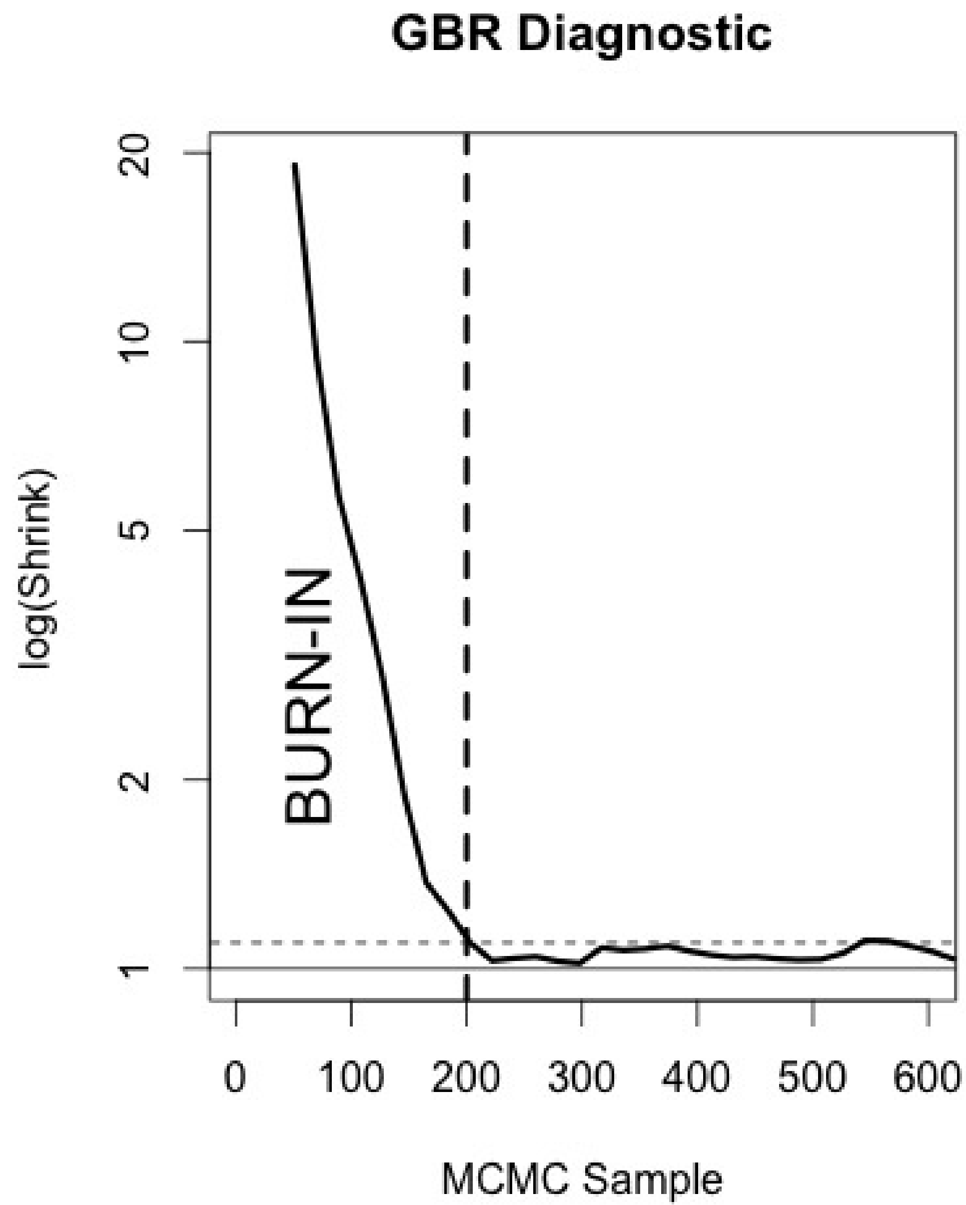
Convergence Statistics

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

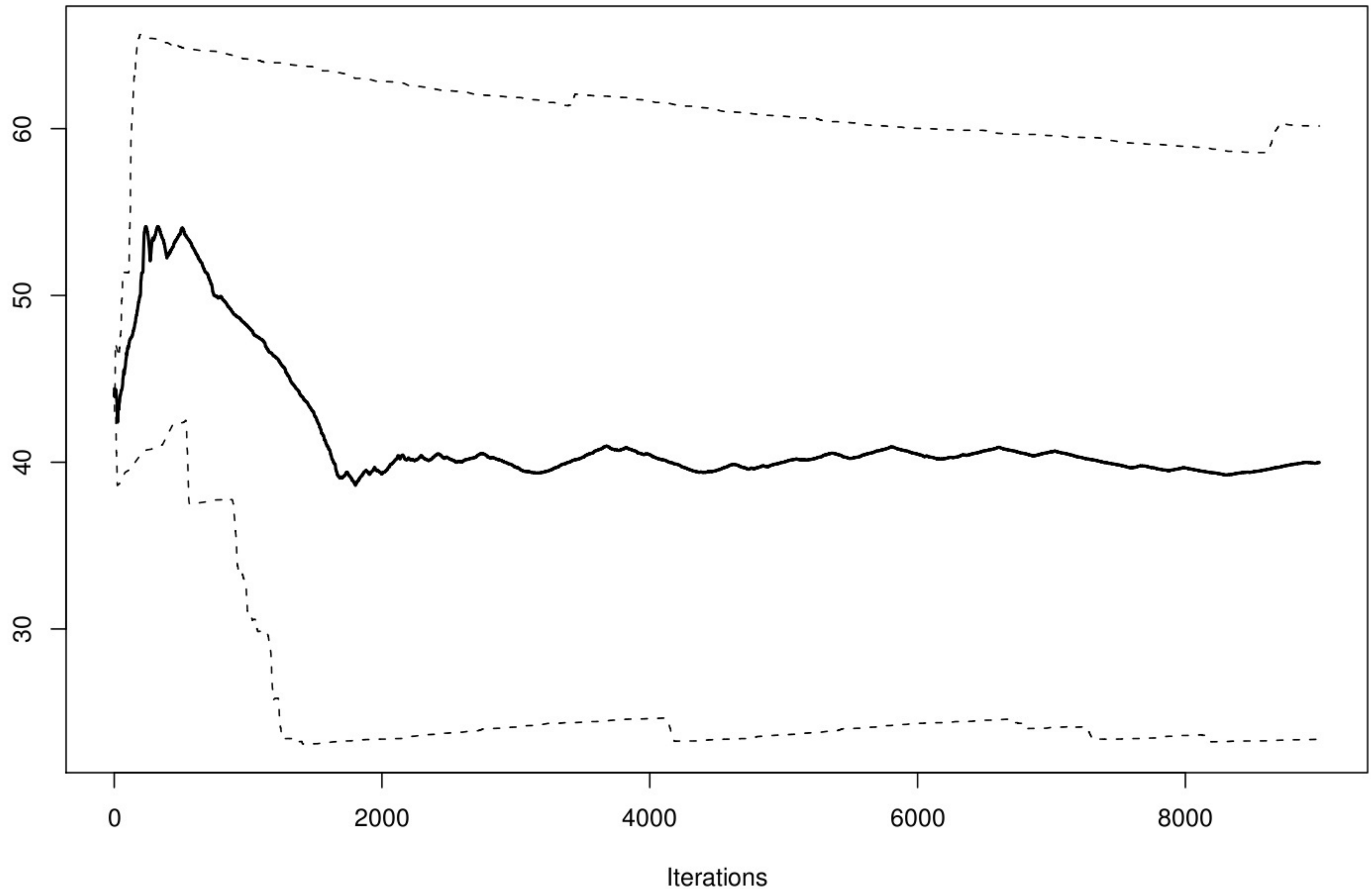


Convergence Statistics

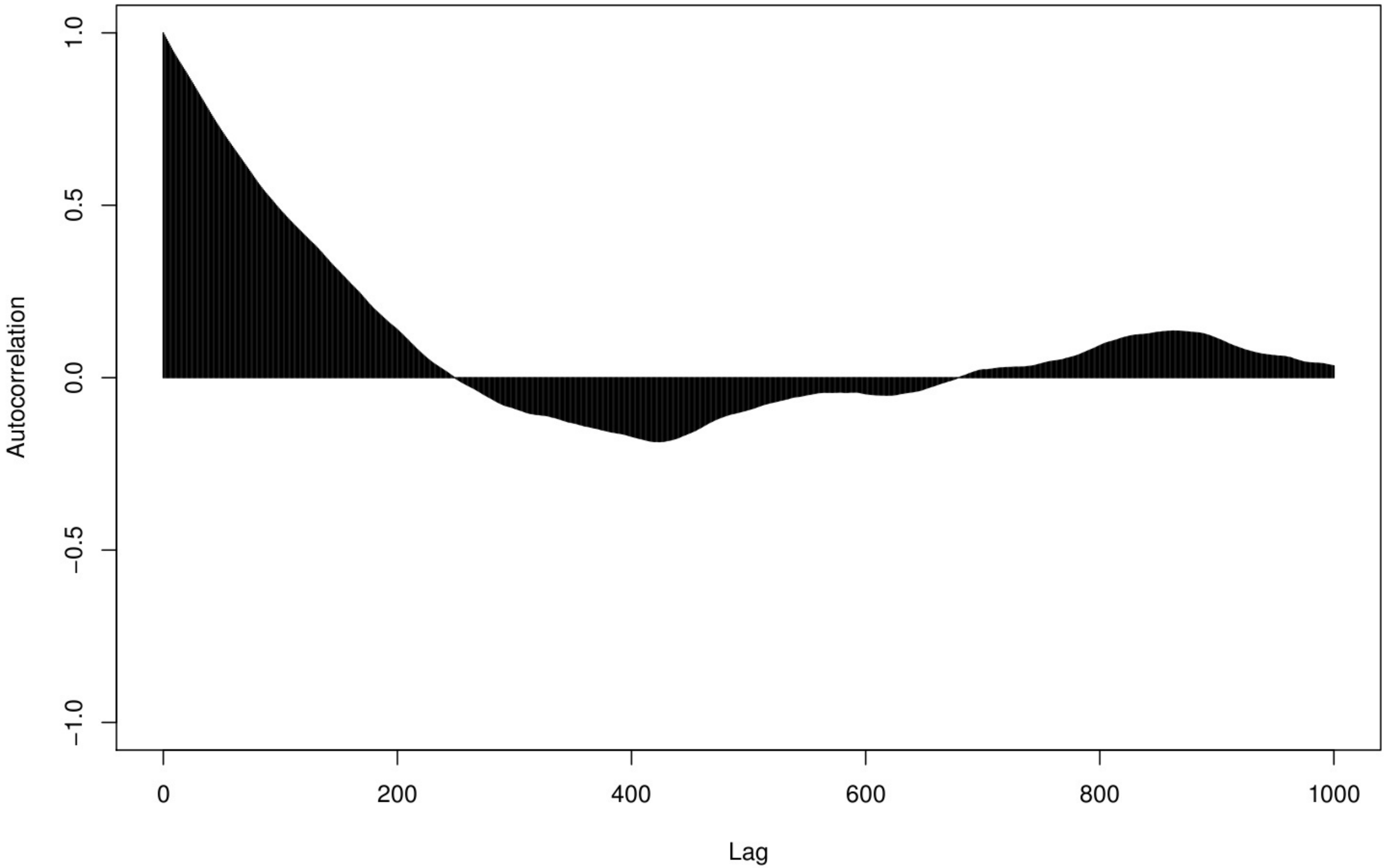




Quantiles



Autocorrelation



Acceptance Rate

- Metropolis & Metropolis – Hastings
  - Aim for 30-70%
  - Too low = not mixing
  - Too high = small steps, slow mixing
  - Example: 97%
- Gibbs sampling
  - Always 100%

Summary Statistics

Analytical:

Mean	SD
43.09901	9.95037

MCMC:

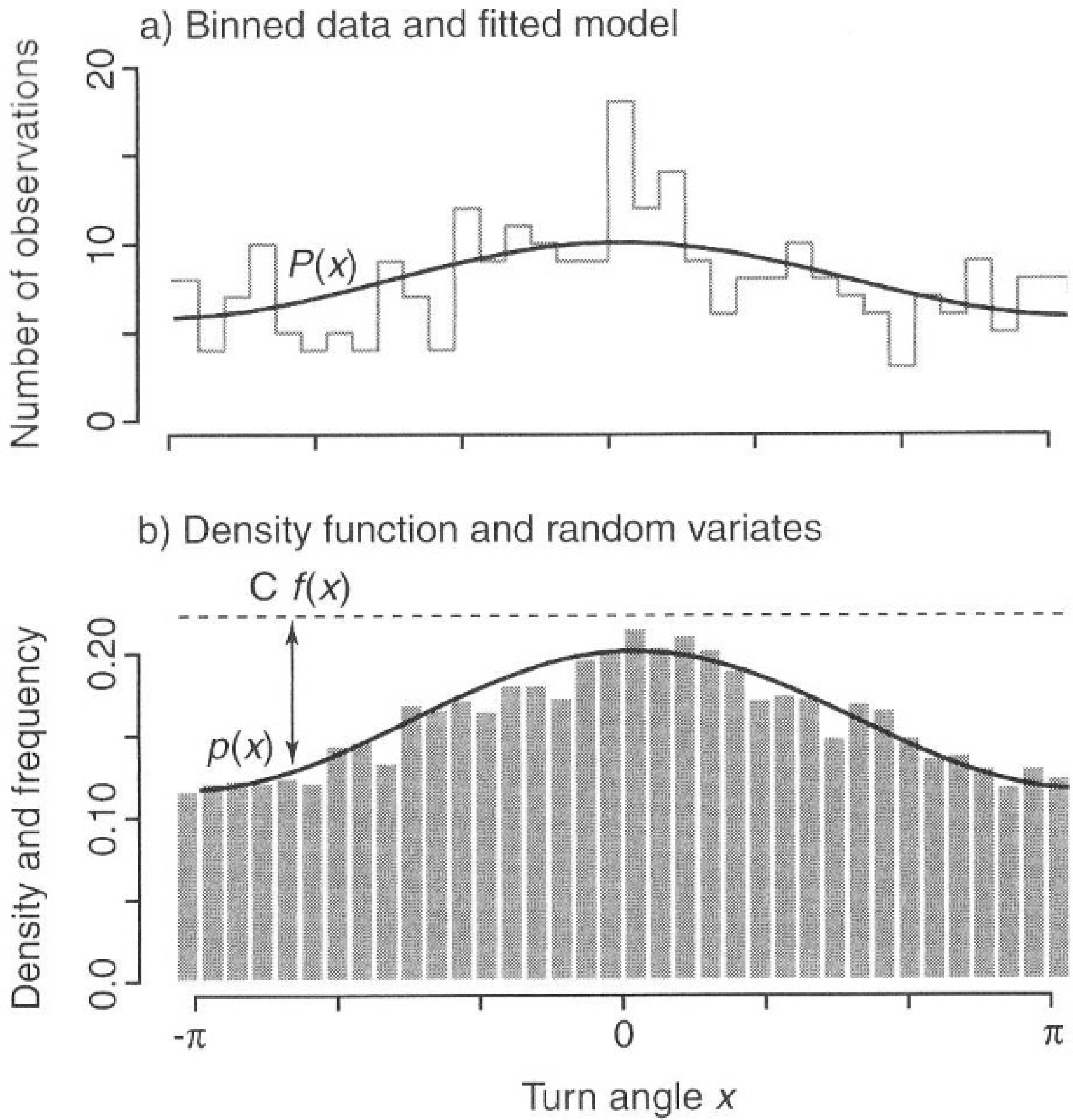
Mean	SD	Naive SE	Time-series
43.05504	9.28108	0.05648	0.74503

Quantiles:

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

Rejection Sampling

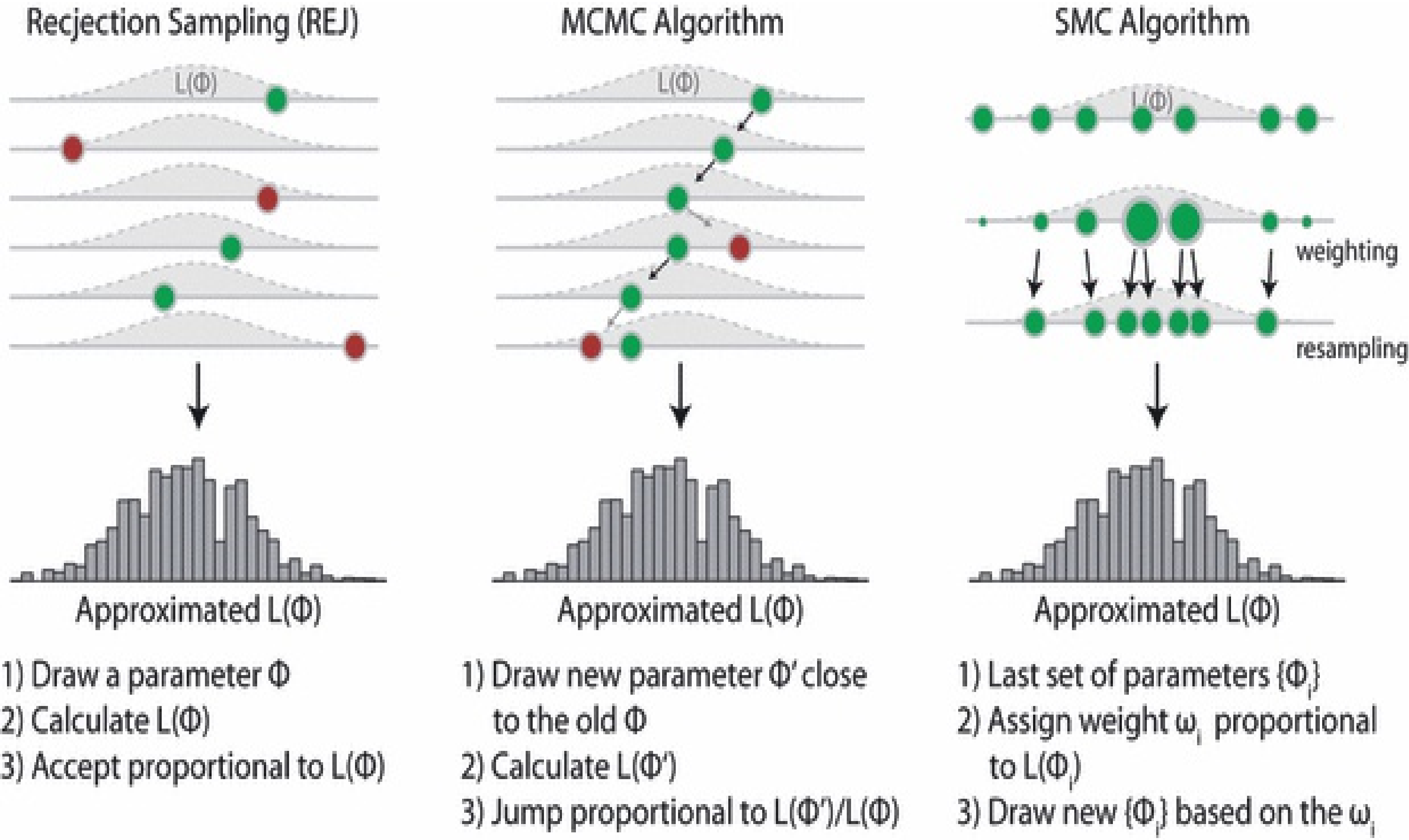
- Want to sample from some distribution g(x)
- Requires that we can sample from a second distribution f(x) such that C*f(x) > g(x) for all x
- Algorithm
  - Draw a random value from f(x)
  - Calculate the density g(x) and f(x) at that x
  - Calculate a = g(x)/[C*f(x)]
  - Accept the proposed x with probability *a* based on a Bernoulli trial
  - If rejected, repeat by proposing a new x...



a) Binned data and fitted model
Number of observations
20
10
0
P(x)
b) Density function and random variates
C f(x)
p(x)
Density and frequency
0.20
0.10
0.0
-π
0
π
Turn angle x

Sequential Monte Carlo (SMC)

- Propose LARGE number of samples from prior
- Calculate Likelihood at each, L_i
- Approximate normalizing constant $P(Y) \alpha \Sigma L_i$
- Calculate weights $w = L_i/P(Y)$
- Resample proportional to weights (Inv CDF)
- Risks:
  - If n is small, weights concentrated
  - Harder in higher dimensions, broad priors
- Through time = Particle Filter



Hartig et al 2011 Ecology Letters