Markov Chain Monte Carlo

Menu

- Motivation, Bayesian Inference
- Metropolis-Hastings algorithm
- Gibbs Sampler
- Special Topics
    - Reversible Jump/ Stochastic Search Variable Selection
    - Particle Filter
    - Parameter Identifiability

Motivation

Dongarra and Sullivan (2000) Guest Editors' Introduction: The Top 10 Algorithms, *Computing in Science and Engineering*, **2**, 22-23.

The top 10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century (in chronological order):

1. Metropolis Algorithm for Monte Carlo
2. Simplex Method for Linear Programming
3. Krylov Subspace Iteration Methods
4. The Decompositional Approach to Matrix Computations
5. The Fortran Optimizing Compiler
6. QR Algorithm for Computing Eigenvalues
7. Quicksort Algorithm for Sorting
8. Fast Fourier Transform
9. Integer Relation Detection
10. Fast Multipole Method

Motivation

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*
(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Motivation

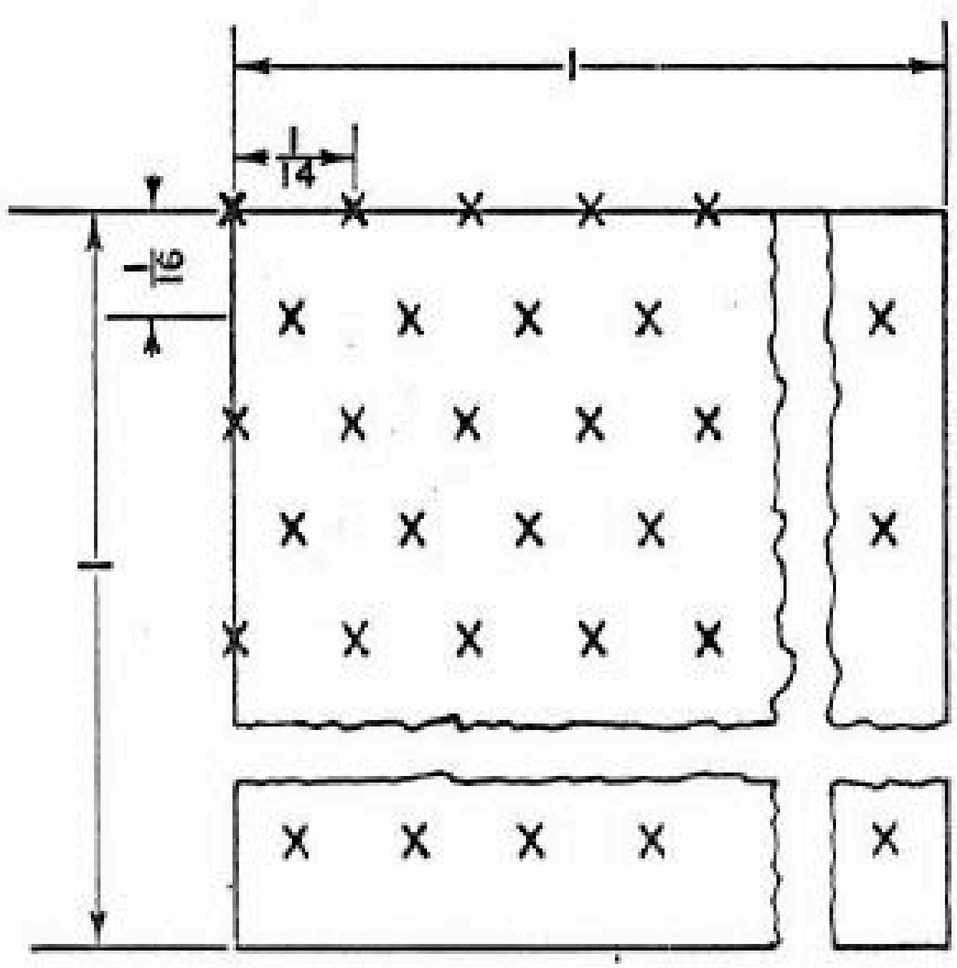


FIG. 2. Initial trigonal lattice.

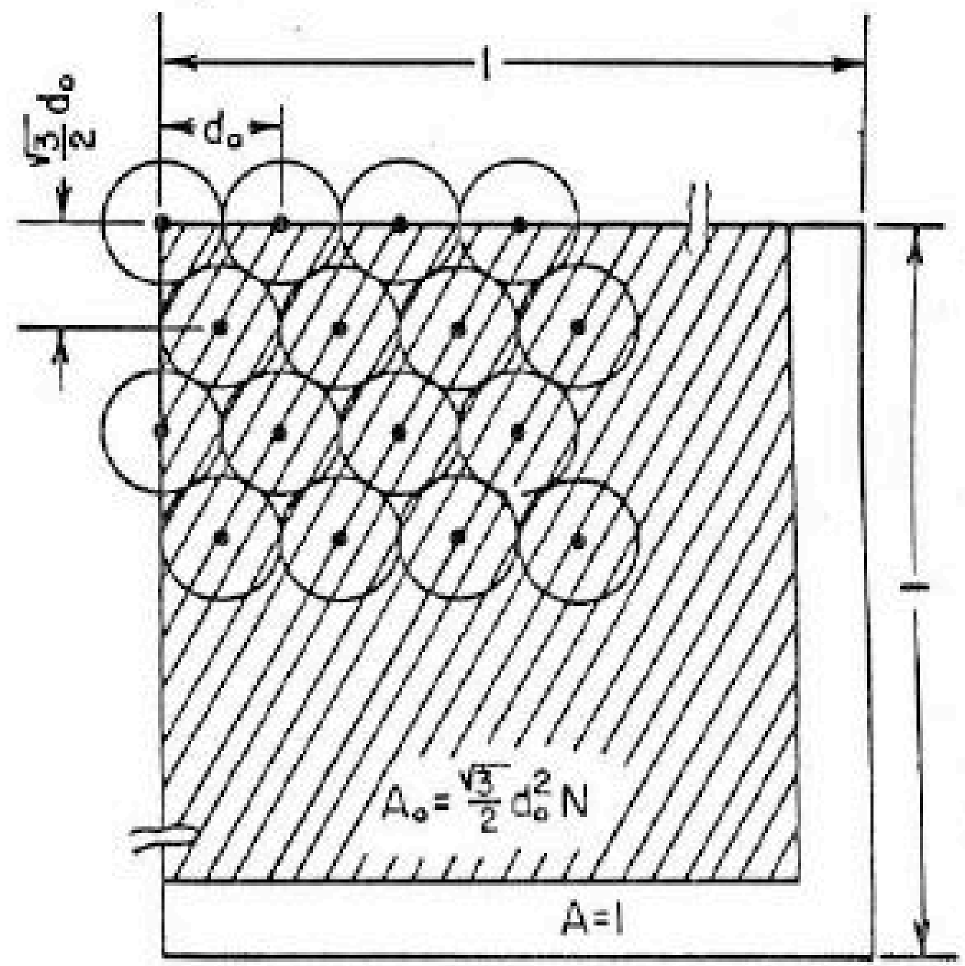


FIG. 3. The close-packed arrangement for determining A_0.

Inference

I've got some data **y** and a model with some parameters θ, what value of θ best explains my observations **y** ?

Frequentist

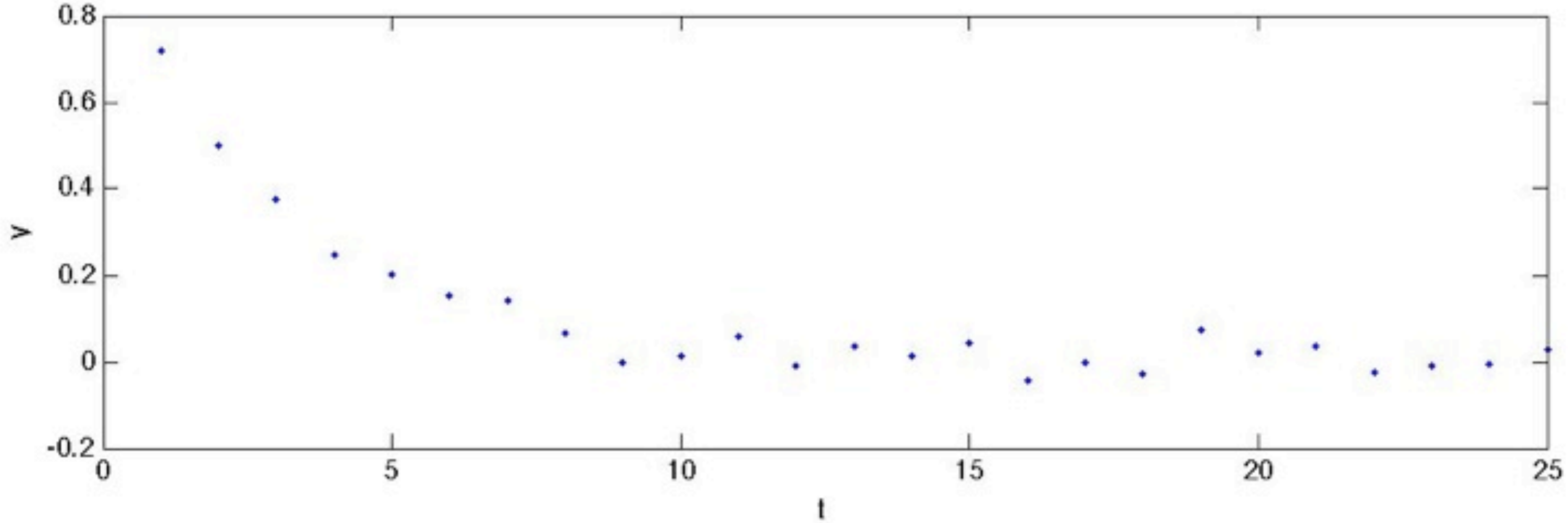
There is some truth out there in the universe, $\underline{\theta}$, which we have only approximate access to, but with infinite sample size and infinite repetition, we could recover $\underline{\theta}$ exactly.

Bayesian

I don't care whether or not there is some ultimate truth. I only have finite data, so how much can I know with my current sample size. That is, what is the full set of possible θ that might explain **y** and what is the probability of each?

Inference

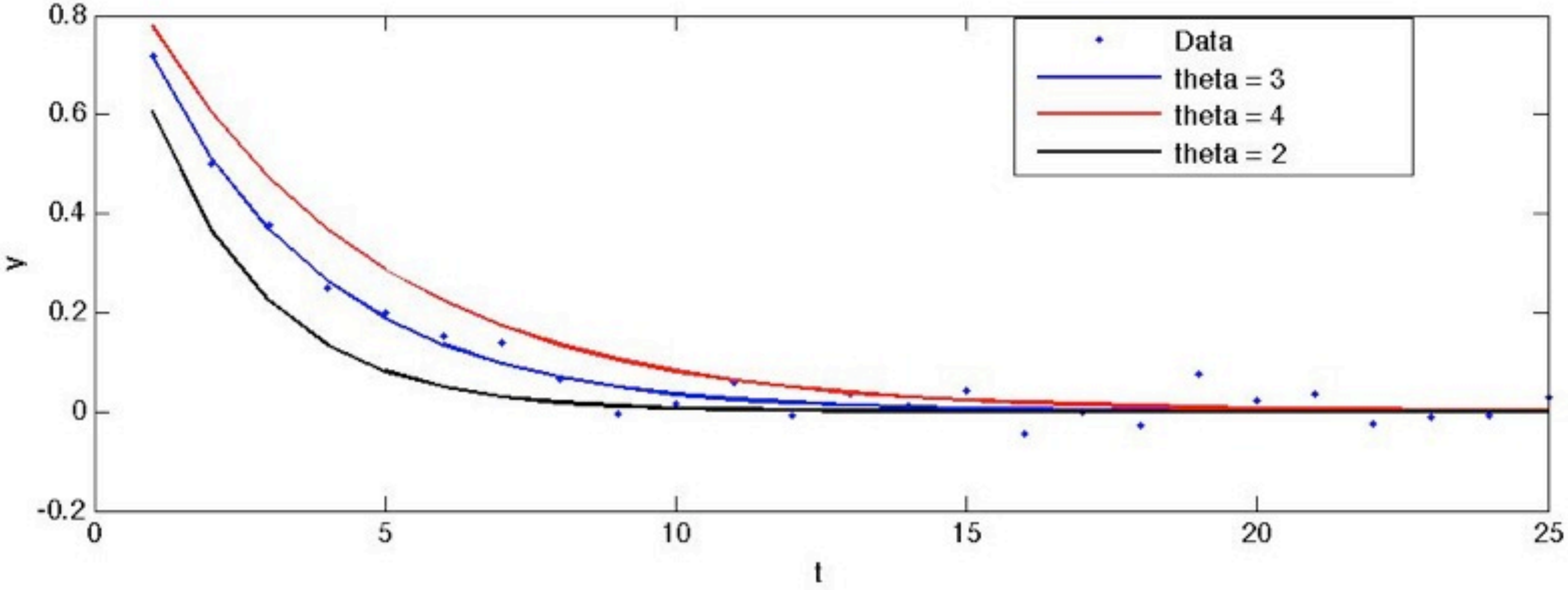
Simple example: curve fitting $y(t) = \exp(-t/\theta)$



Pretend each observed point is drawn iid from a normal distribution with mean $y(t)$ and some variance

Inference

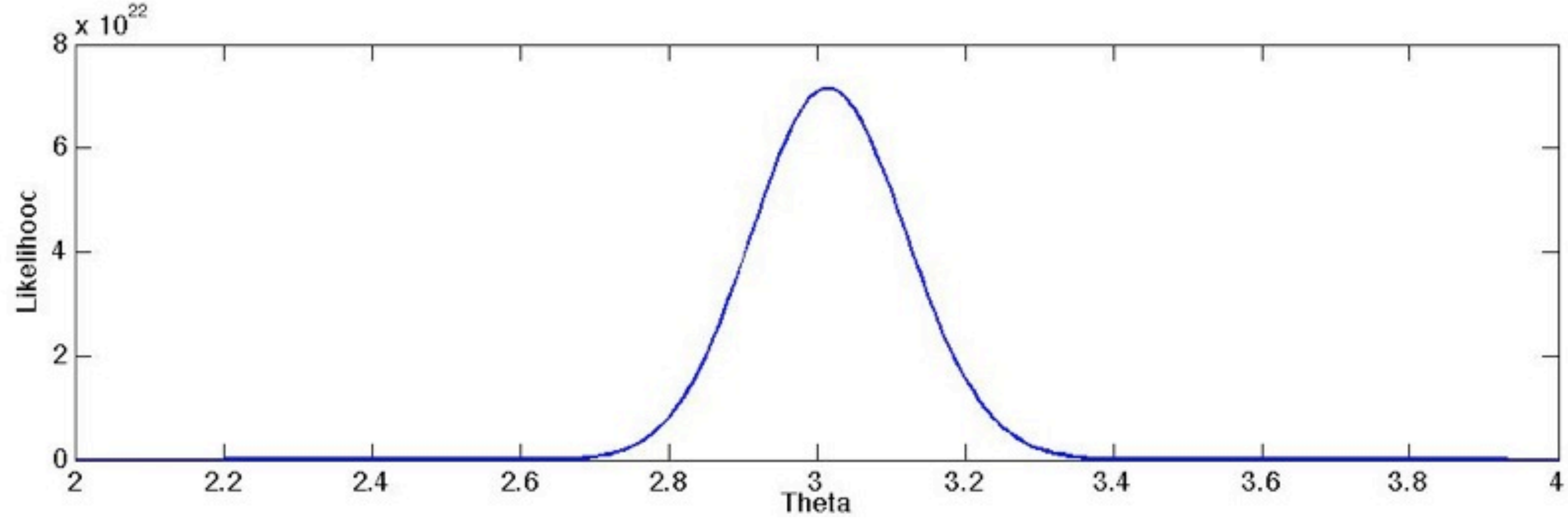
Simple example: curve fitting $y(t) = \exp(-t/\theta)$



We're trying to find θ that explains the observations. Some seem to "fit" better than others.

Inference

Maximum likelihood: $\underline{\theta} = 3$



Generally, MLE provides a point estimate in parameter space with little indication of how sensitive the data is to variation across parameters

Bayes

Instead of a point estimate, I'd rather know the entire distribution $p(\theta|\mathbf{y})$. Using Bayes' rule :

$$\underbrace{p(\theta|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\theta)}_{\text{Likelihood}}\underbrace{p(\theta)}_{\text{Prior}}$$

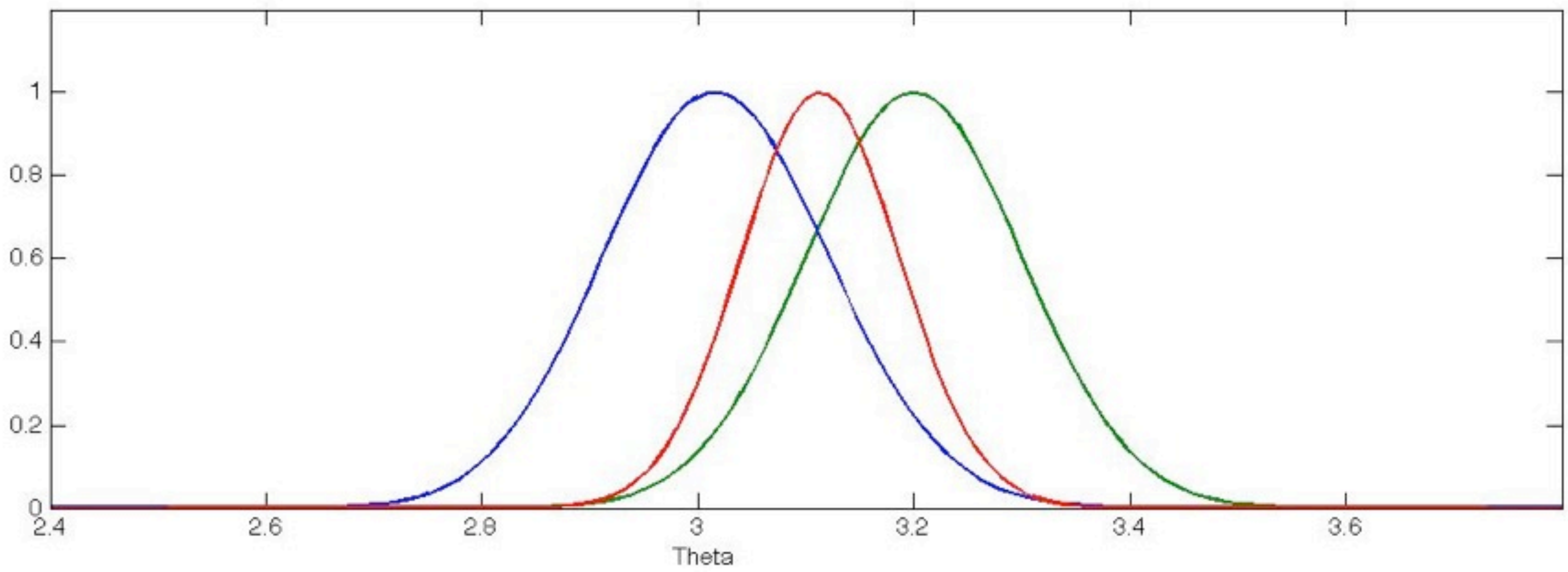
Bayes

$$\underline{p(\theta|\mathbf{y})} \propto \underline{p(\mathbf{y}|\theta)}\,\underline{p(\theta)}$$

Posterior

Likelihood

Prior



Bayes - Summary

I have observations **y** and a model with parameters θ, what the is posterior distribution over θ given **y**?

The following MCMC methods are often used to sample from posterior distributions from inference problems, but can be used to sample from any distributions of interest.

Monte Carlo Integration

Replace exact integral with approximate sum

$$\int f(x)p(x)dx \approx \sum_{i=1}^{N} f(x_i)$$

if x_i are drawn iid from p(x)

So how to draw iid samples from arbitrary p(x)?

Metropolis-Hastings

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Metropolis-Hastings

Construct a Markov chain $\{\theta_0, \theta_1, \theta_2, ...\}$ in the following way:

0. Start with initial θ_0 arbitrarily

1. Generate a proposal move $\tilde{\theta} \sim q(\tilde{\theta}|\theta)$

2. Accept $\tilde{\theta}$ with probability $\alpha = min\left(1, \frac{p(\tilde{\theta})}{p(\theta)}\frac{q(\theta|\tilde{\theta})}{q(\tilde{\theta}|\theta)}\right)$

Metropolis Random Walk

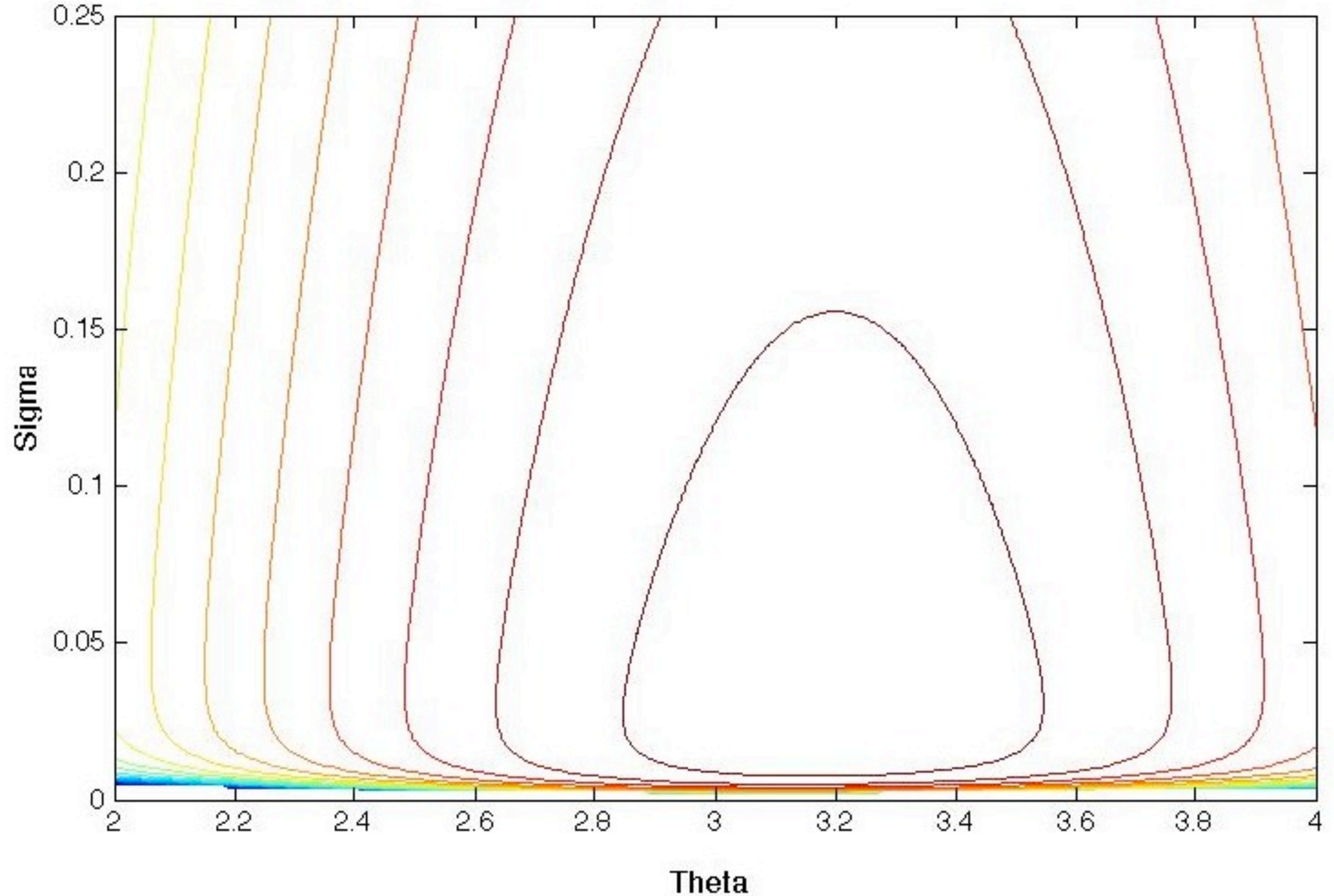
Construct a Markov chain $\{\theta_0, \theta_1, \theta_2, ...\}$ in the following way:

0. Start with initial θ_0 arbitrarily

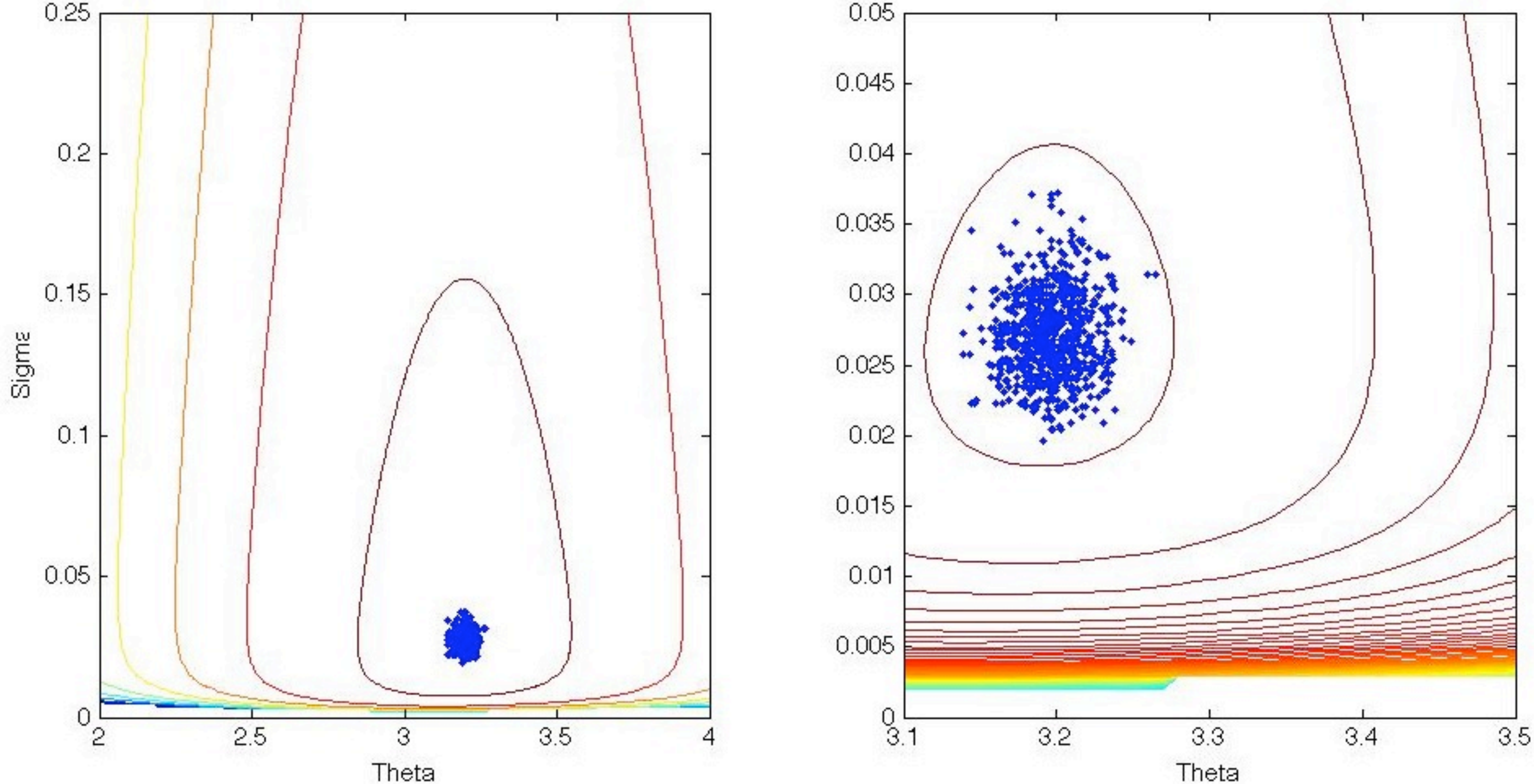
1. Generate a proposal move $\tilde{\theta} \sim N(\theta, \sigma)$

2. Accept $\tilde{\theta}$ with probability $\alpha = min\left(1, \frac{p(\tilde{\theta})}{p(\theta)}\right)$

Metropolis Random Walk

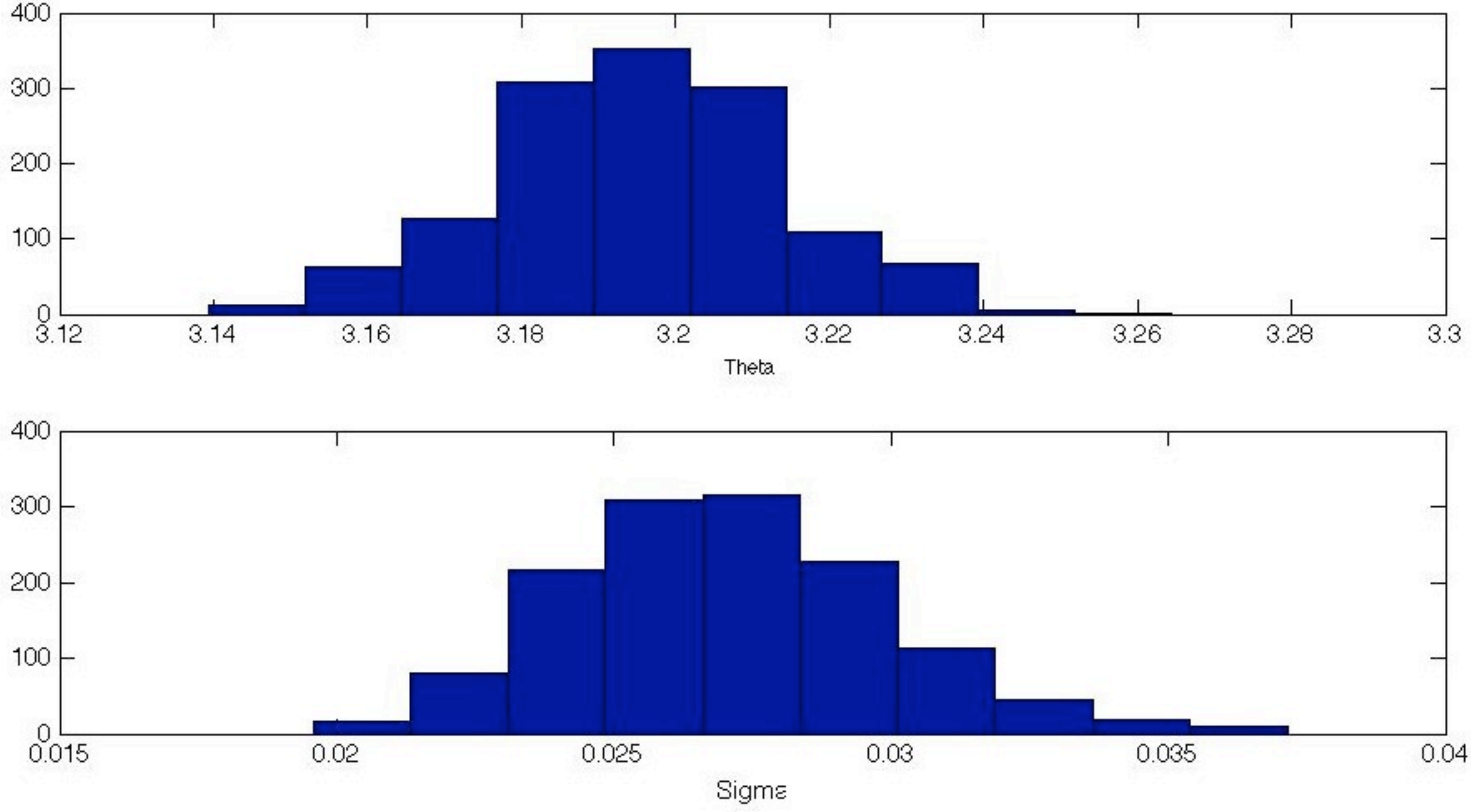


Metropolis Random Walk

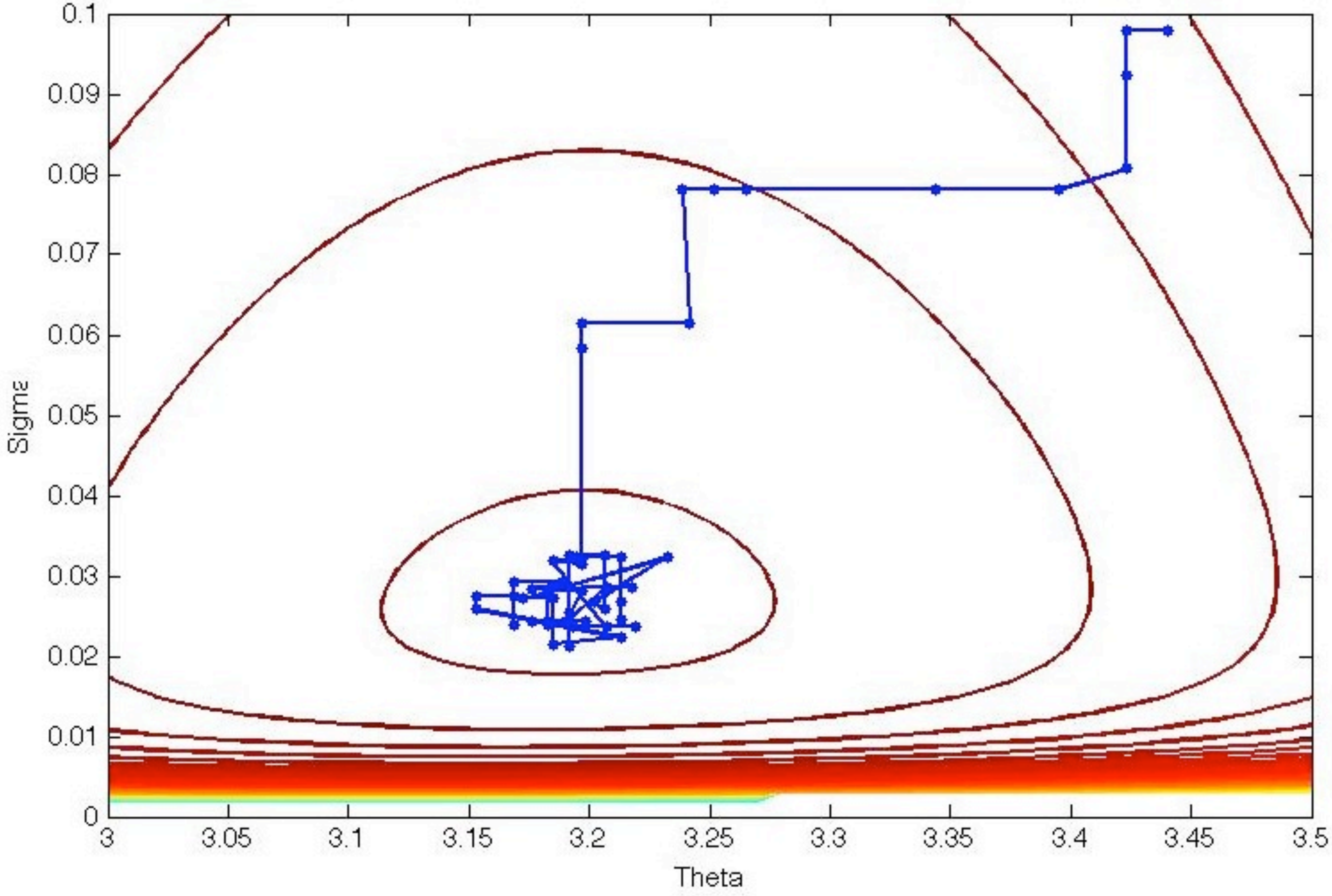


Metropolis Random Walk

Marginal Posterior Distributions

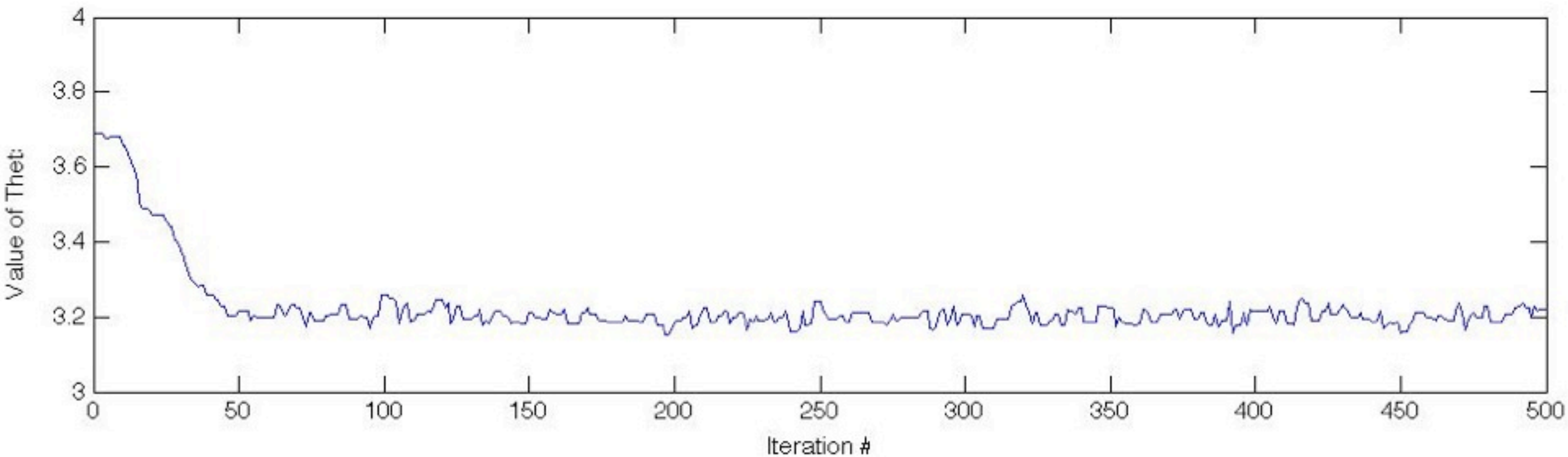


Metropolis Random Walk



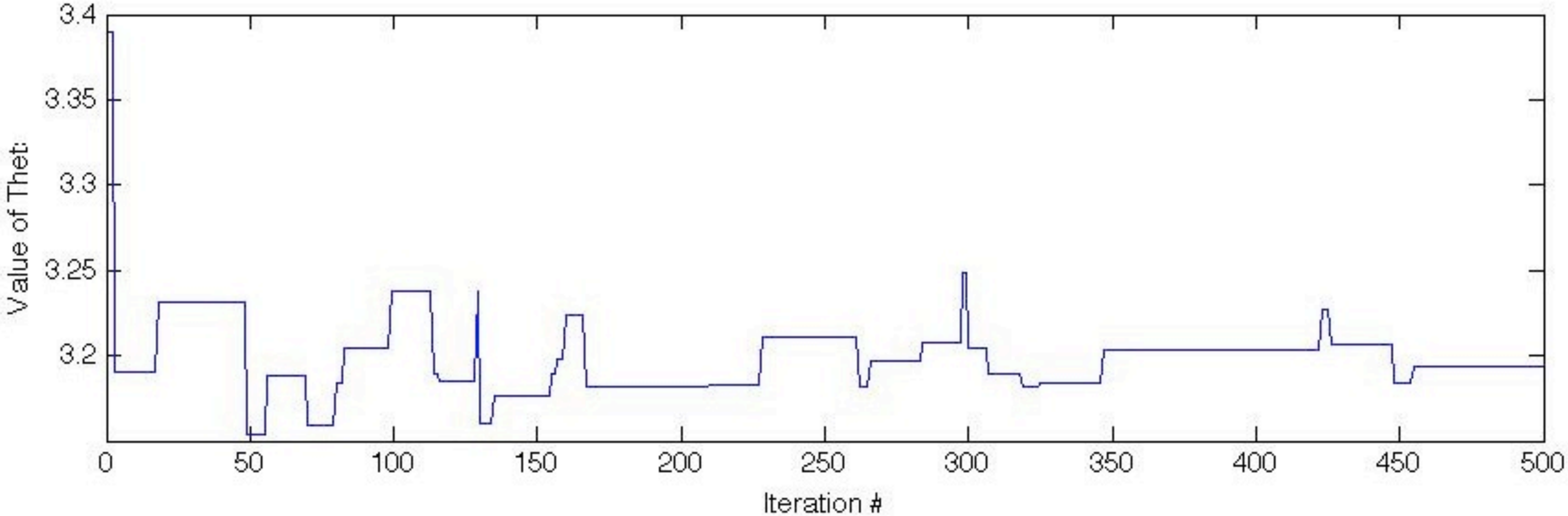
Metropolis Random Walk

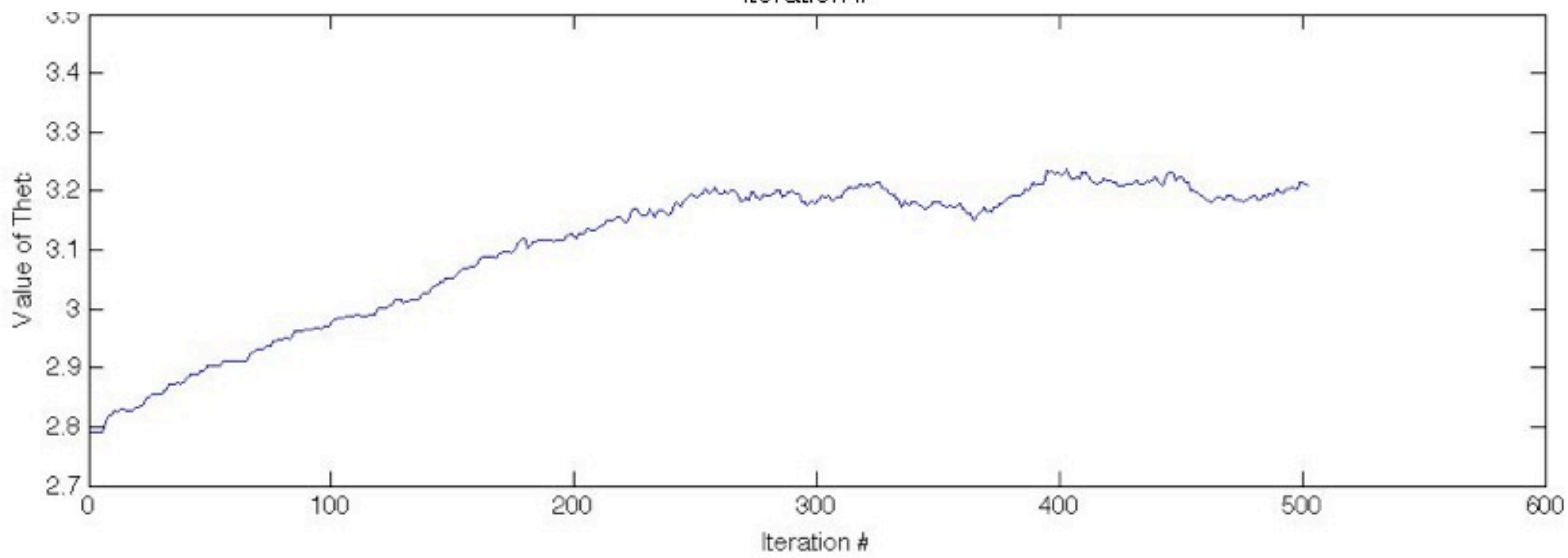
Convergence / Mixing



Metropolis Random Walk

Convergence / Mixing



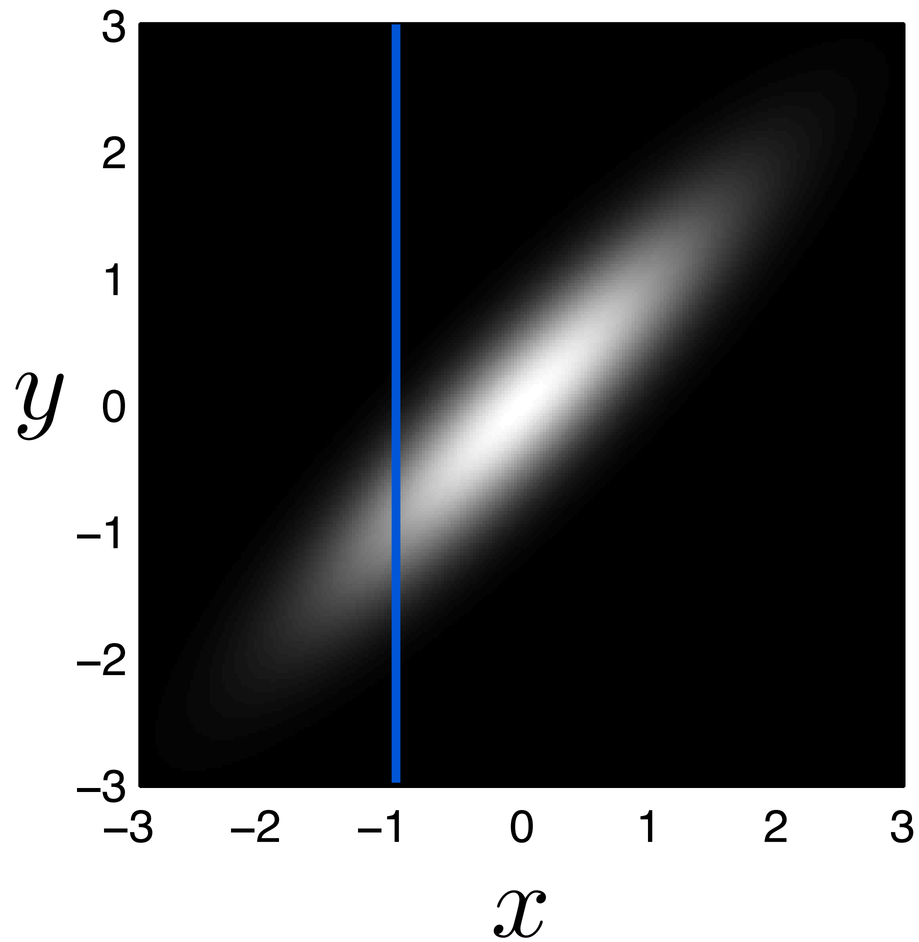


Gibbs Sampler

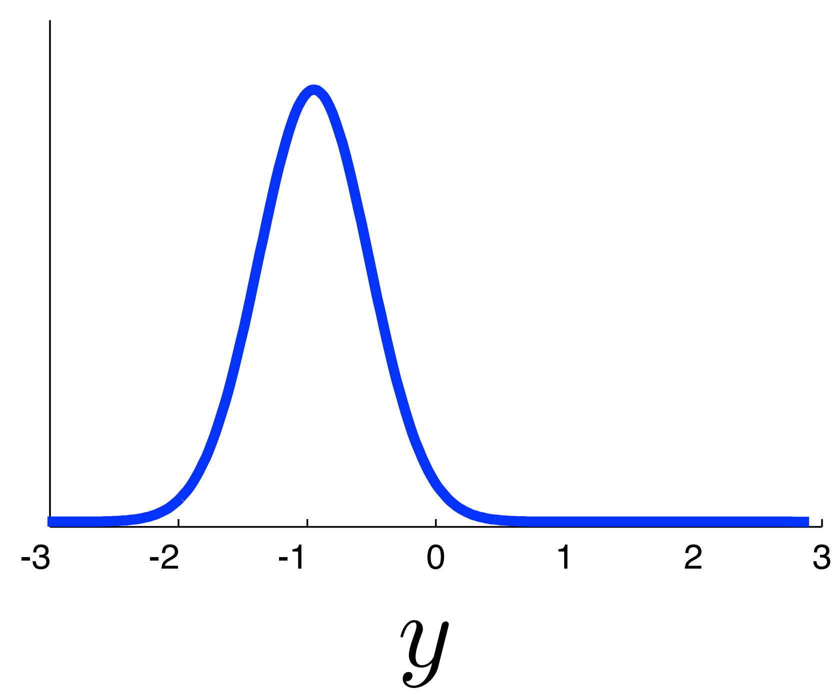
Gibbs Sampler

Conditional Distributions

$$P(x, y)$$



$$P(y|x = -1) = \frac{P(y, x = -1)}{P(x = -1)}$$

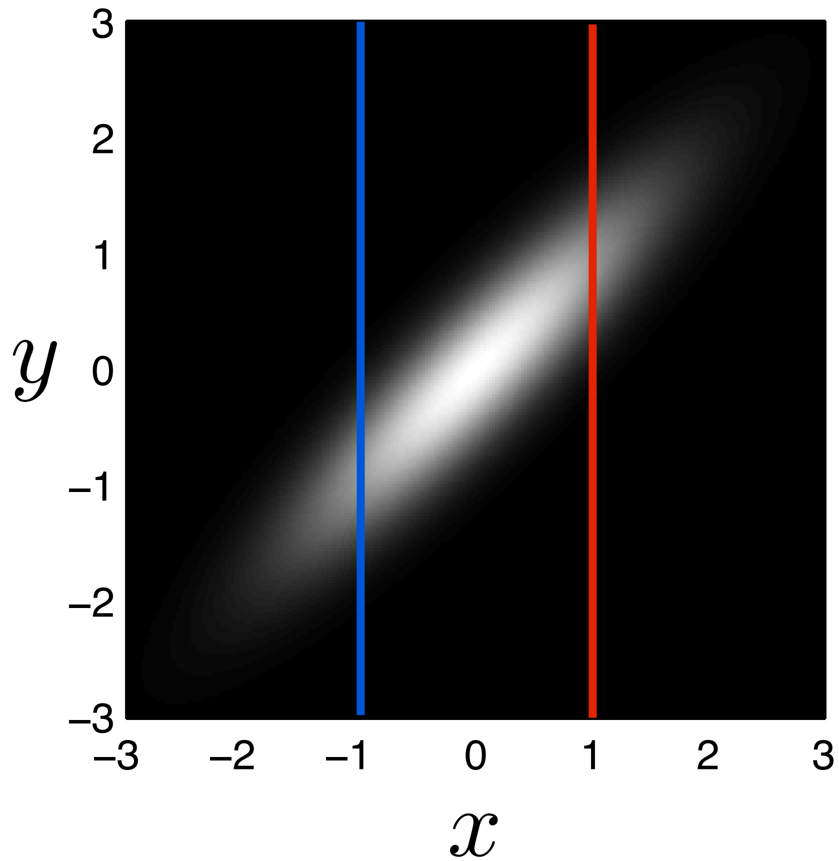


(Imagery stolen from Jonathan Pillow)

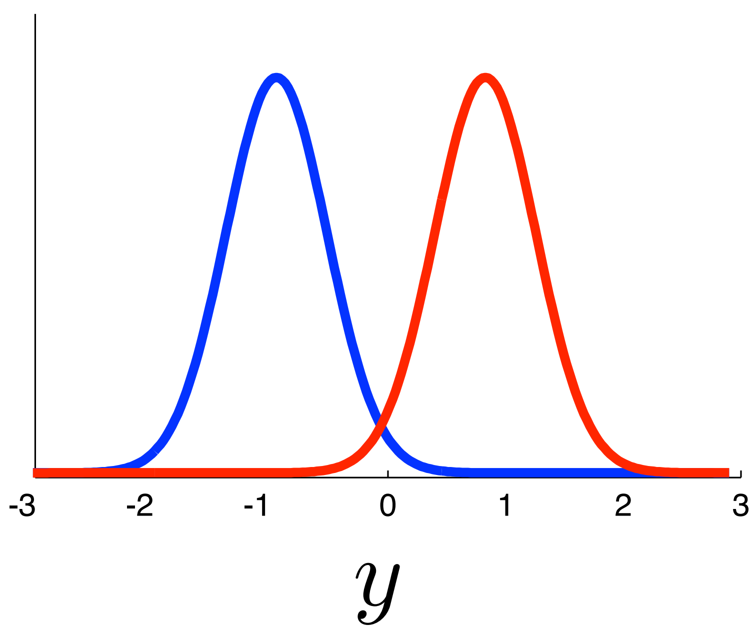
Gibbs Sampler

Conditional Distributions

$$P(x, y)$$



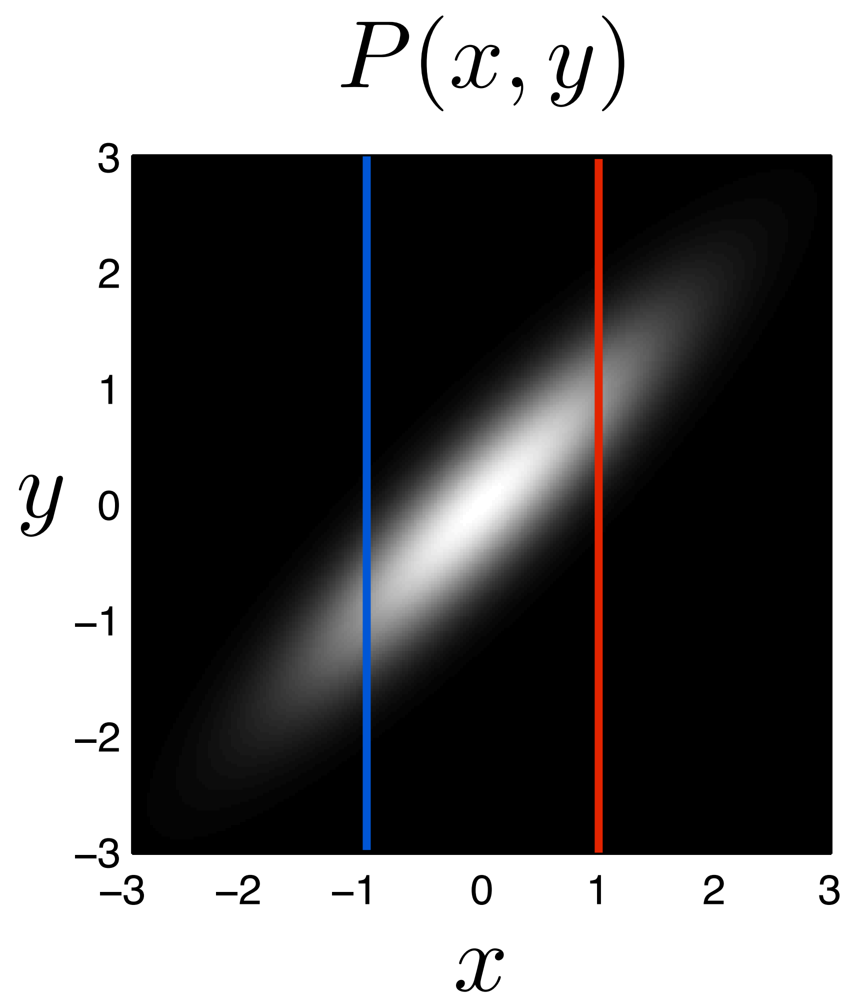
$$P(y|x = 1) = \frac{P(y, x = 1)}{P(x = 1)}$$



(Imagery stolen from Jonathan Pillow)

Gibbs Sampler

Conditional Distributions



$$p(y|x=1) = \frac{p(y, x=1)}{p(x=1)}$$

$$p(y|x=1) \propto p(y, x=1)$$

More generally

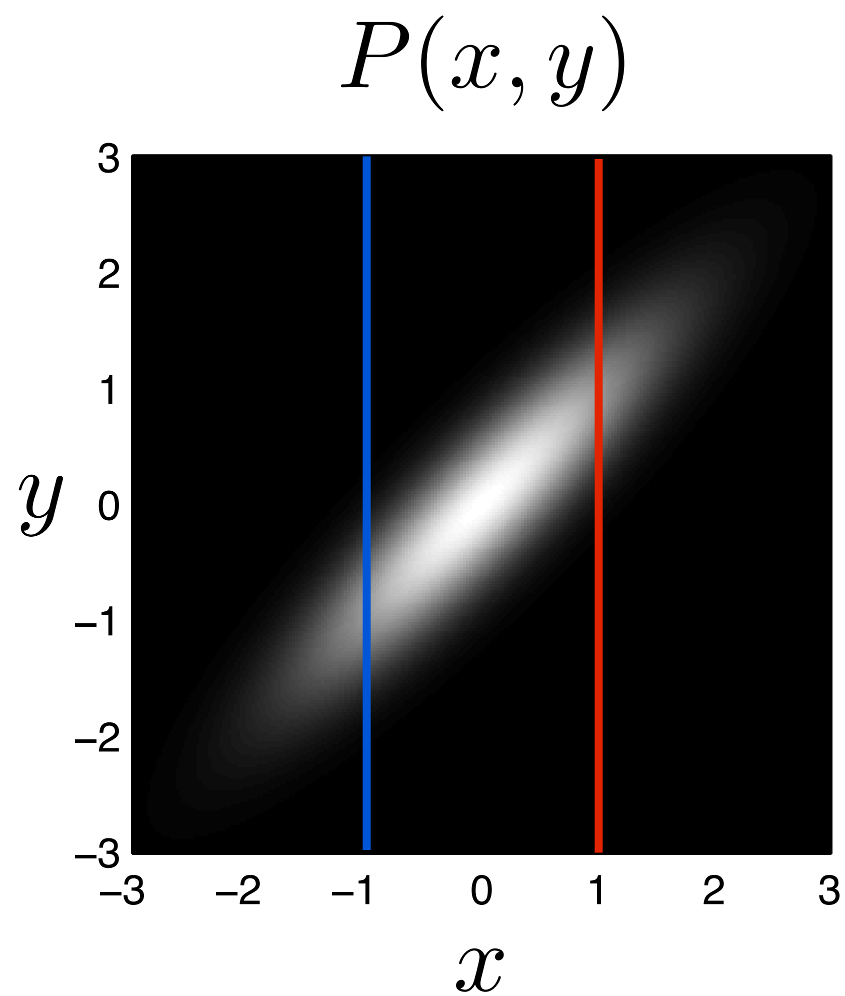
$$p(x,y) \propto p(x|y) \text{ or } p(y|x)$$

A joint probability will be linearly proportional to any of its conditional probabilities

(Imagery stolen from Jonathan Pillow)

Gibbs Sampler

Conditional Distributions



$p(x,y) \propto p(x|y)$ or $p(y|x)$

We can generate samples from joint probability just by generating sets of samples from all conditional probabilities.

(Imagery stolen from Jonathan Pillow)

Gibbs Sampler

Conditional Distributions

$$
\begin{aligned}
p(\theta, \sigma | Data) &\propto p(Data|\theta, \sigma) p(\theta) p(\sigma) \\
&\propto N(Data; \mu, \sigma) N(\theta; a, b) B(\sigma; c, d) \\
p(\theta | Data, \sigma = \sigma_o) &\propto N(Data; \mu, \sigma = \sigma_o) N(\theta; a, b) B(\sigma = \sigma_o; c, d) \\
p(\sigma | Data, \theta = \theta_o) &\propto N(Data; \mu, \sigma) N(\theta = \theta_o; a, b) B(\sigma; c, d)
\end{aligned}
$$

Gibbs Sampler

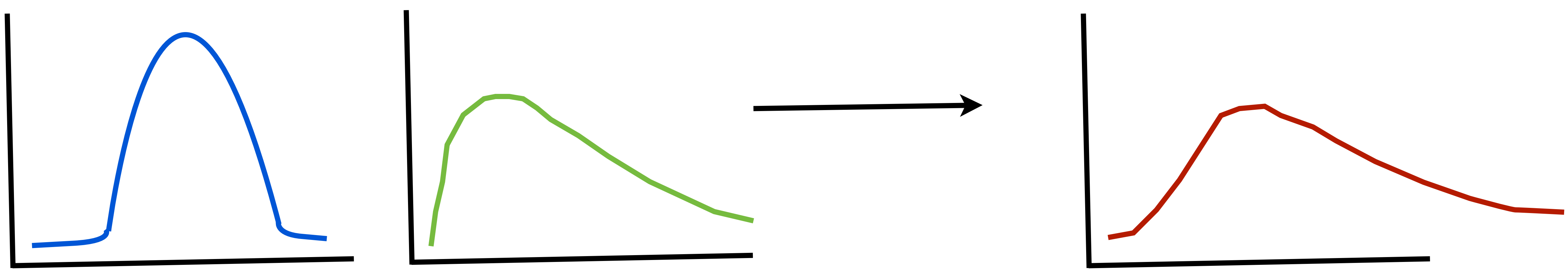
Inverse CDF method

$$p(\sigma|Data, \theta = \theta_o) \propto N(Data; \mu, \sigma)\cancel{N(\theta = \theta_o; a, b)}B(\sigma; c, d)$$

Gibbs Sampler

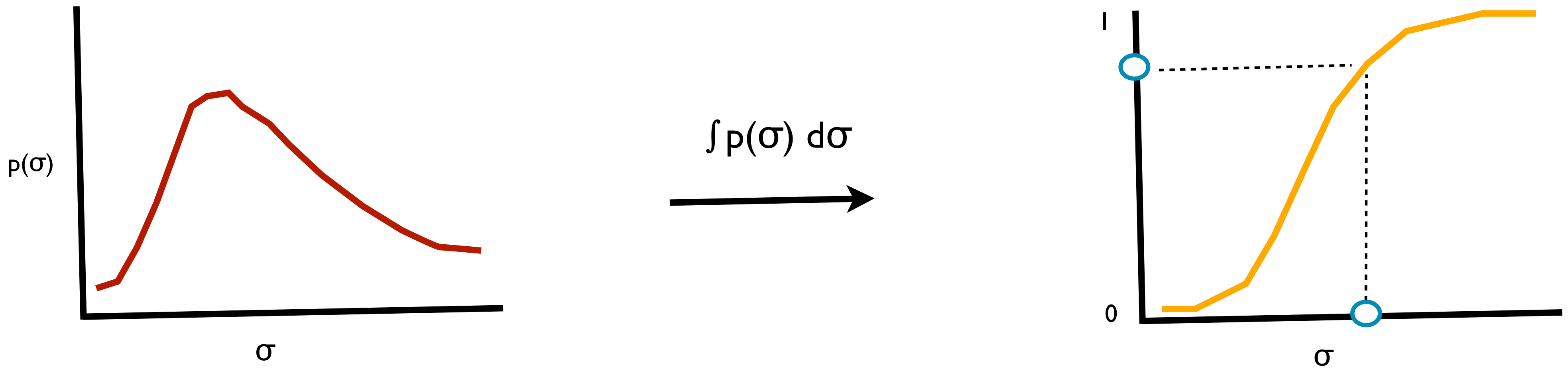
Inverse CDF method

$$p(\sigma|Data, \theta = \theta_o) \propto N(Data; \mu, \sigma)\cancel{N(\theta = \theta_o; a, b)}B(\sigma; c, d)$$

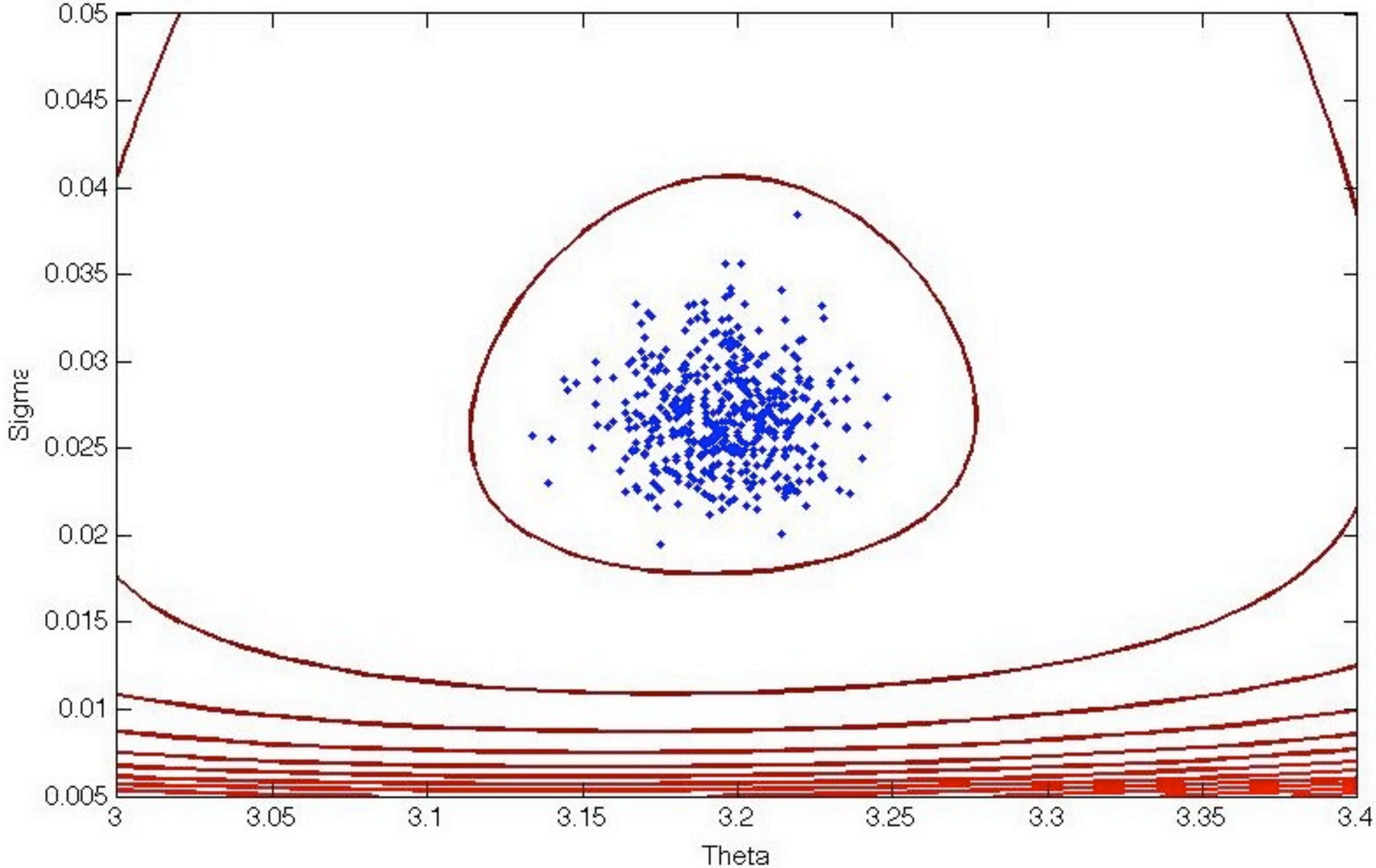


Gibbs Sampler

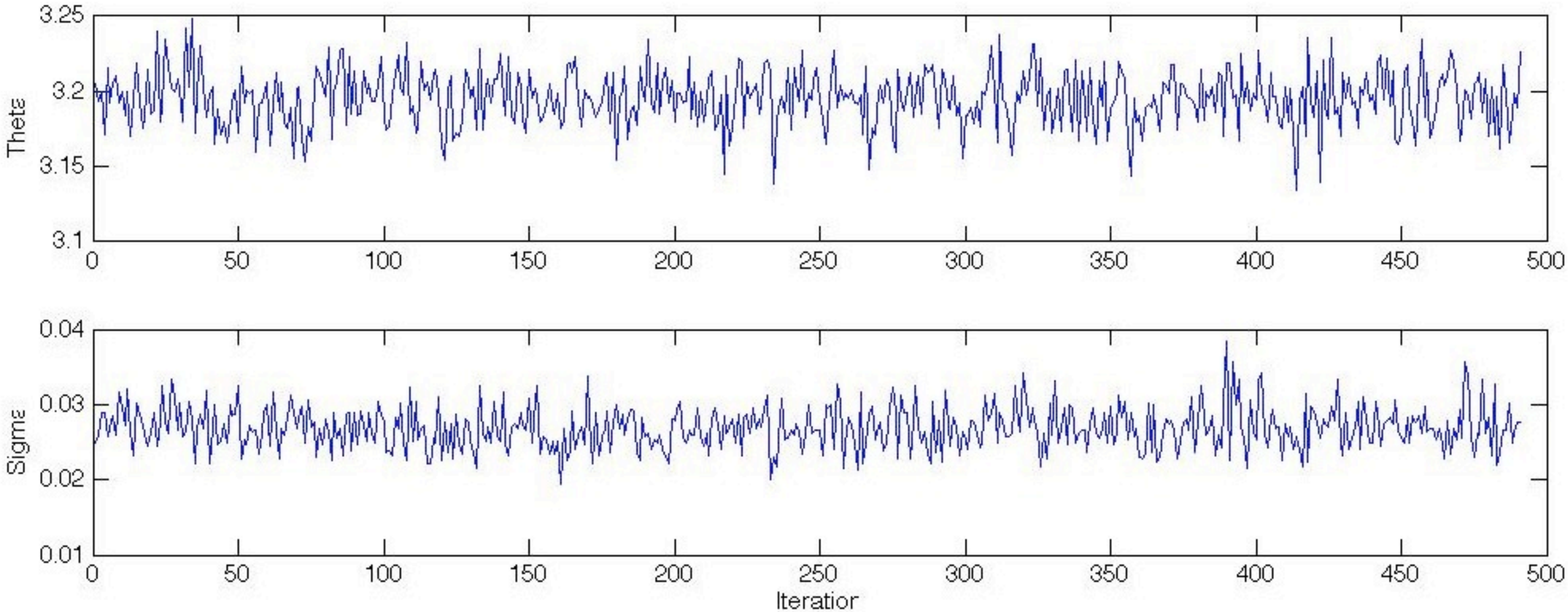
Inverse CDF method



Gibbs Sampler



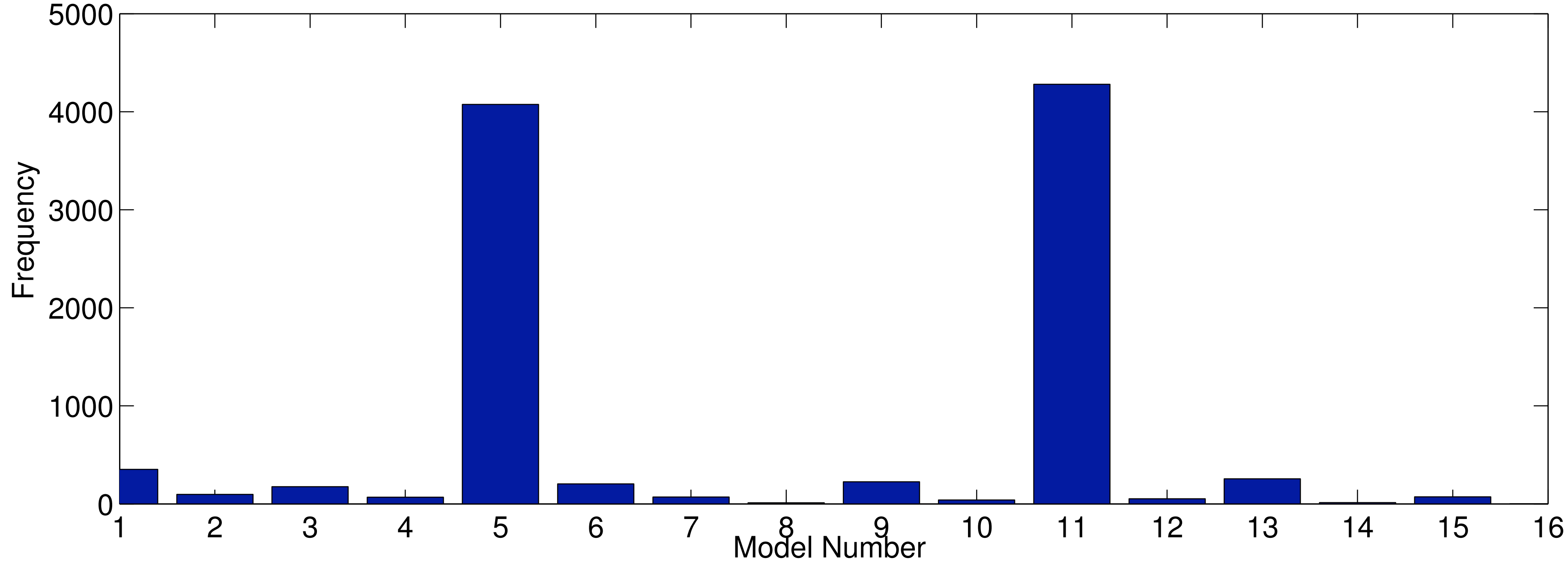
Gibbs Sampler



Special Topics

Model Selection

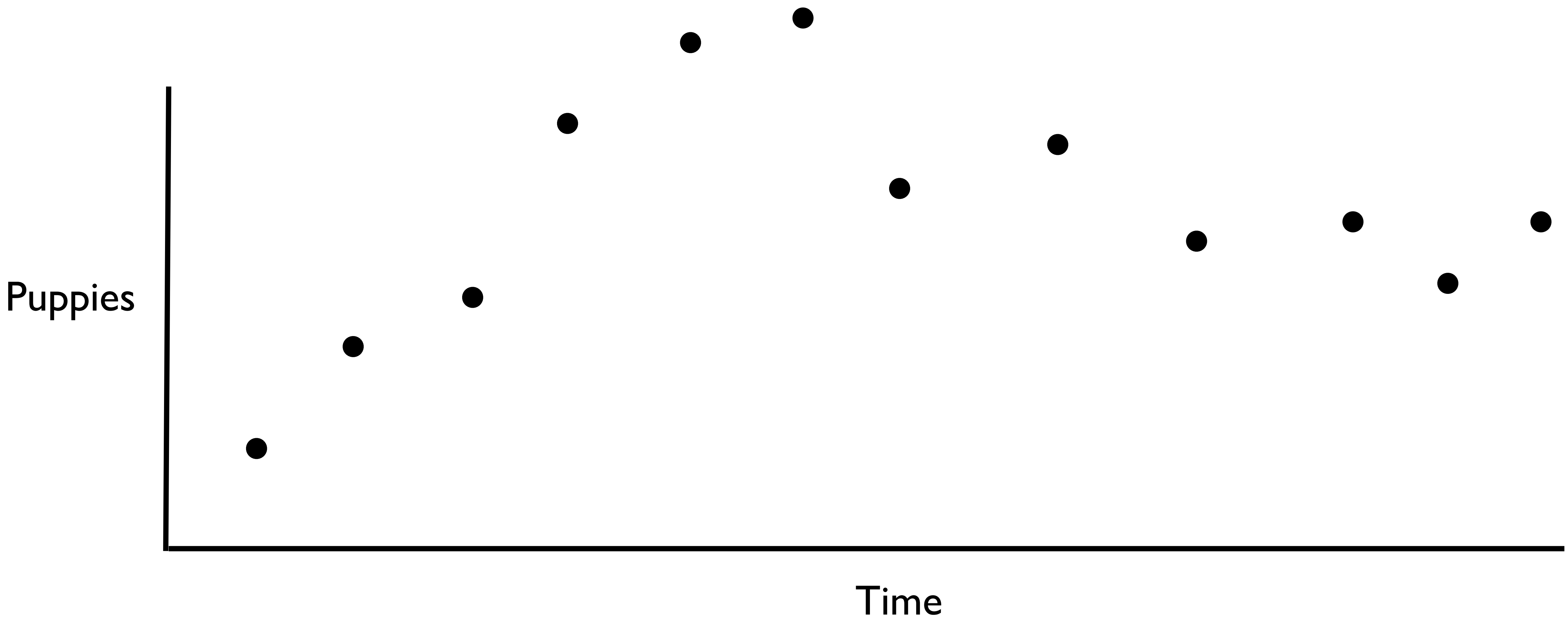
$$p(M|\text{Data},\theta)$$



Reversible Jump/Transdimensional MCMC and Stochastic Search Variable Selection

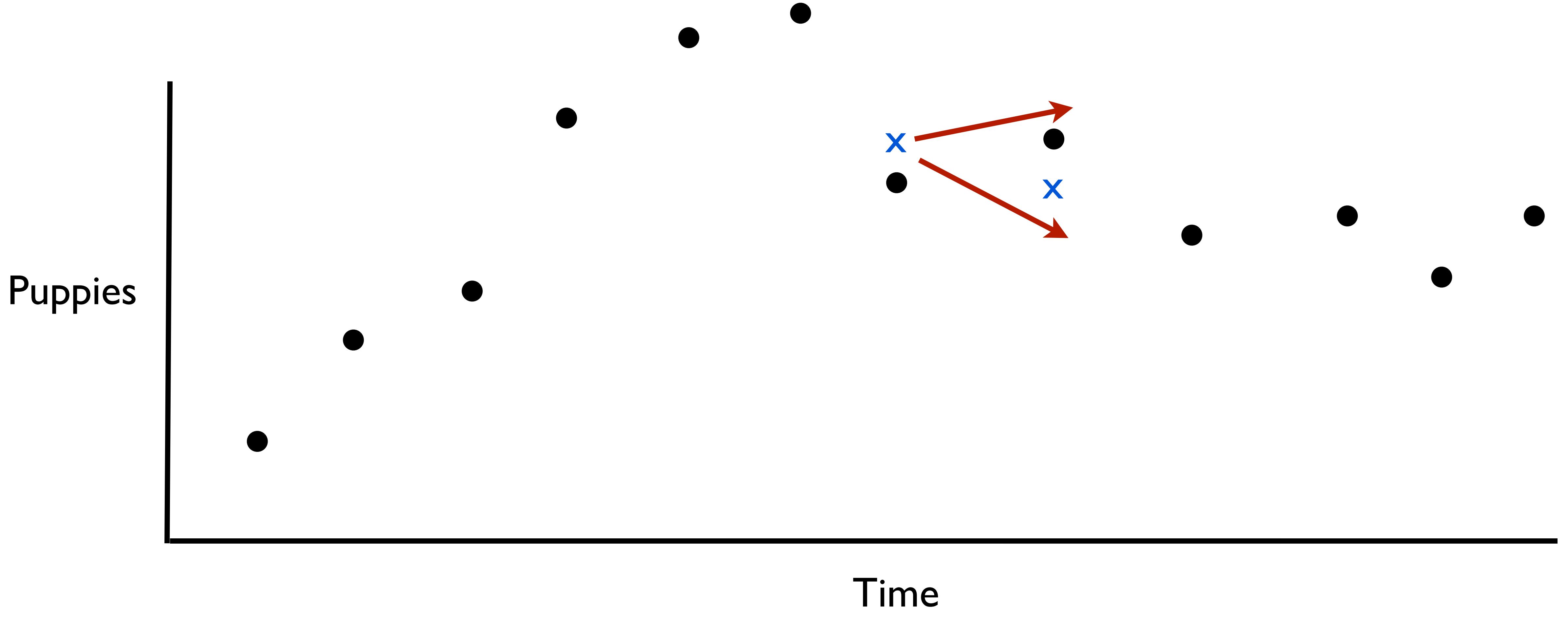
Special Topics

Particle Filter



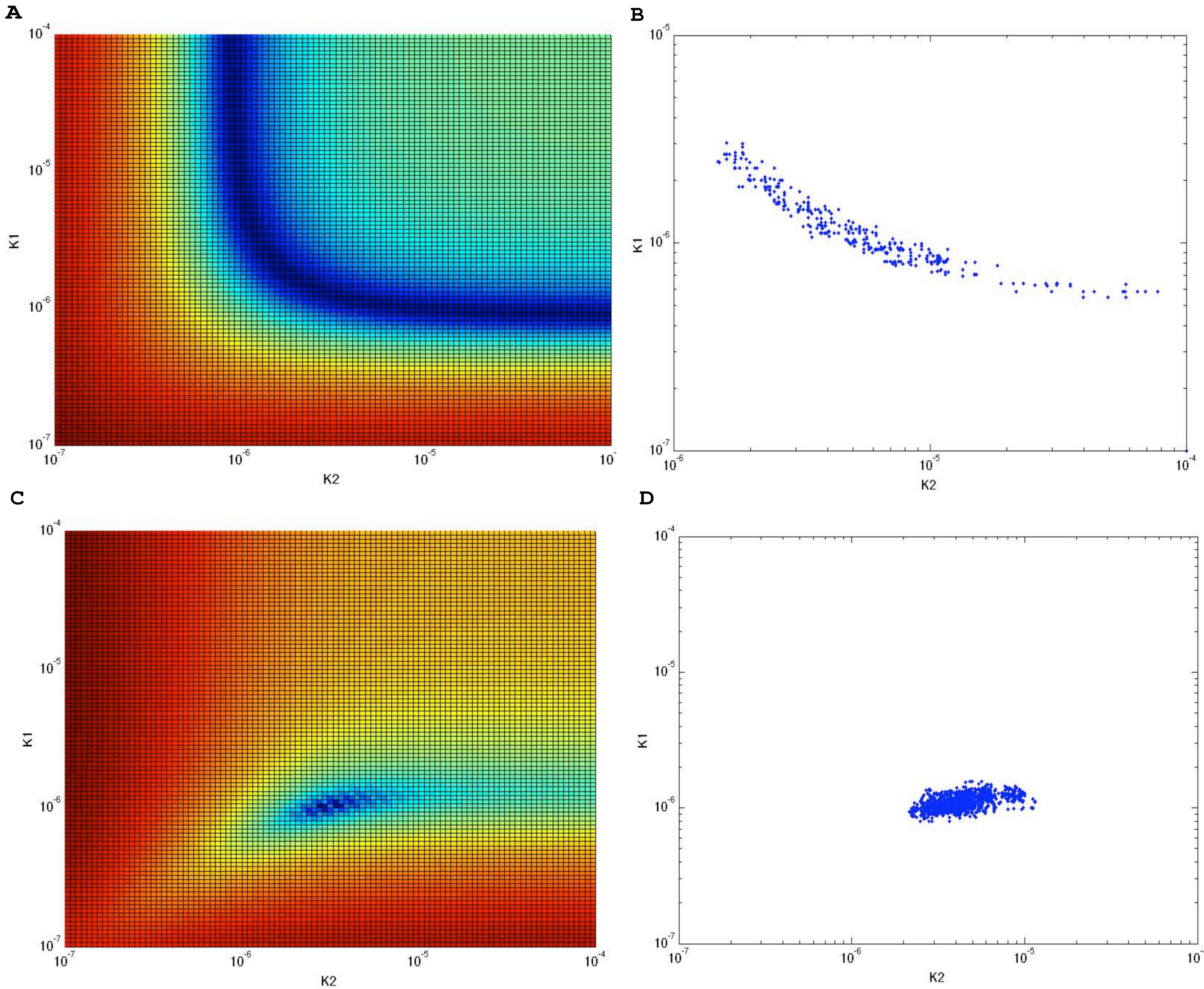
Special Topics

Particle Filter



Special Topics

Parameter Identifiability



Conclusion