Bayesian Techniques for Parameter Estimation

Statistical Inference

Goal: The goal in statistical inference is to make conclusions about a phenomenon based on observed data.

Frequentist: Observations made in the past are analyzed with a specified model. Result is regarded as confidence about state of real world.

- Probabilities defined as frequencies with which an event occurs if experiment is repeated several times.

- Parameter Estimation:

o Relies on estimators derived from different data sets and a specific sampling distribution.

o Parameters may be unknown but are fixed and deterministic.

Bayesian: Interpretation of probability is subjective and can be updated with new data.

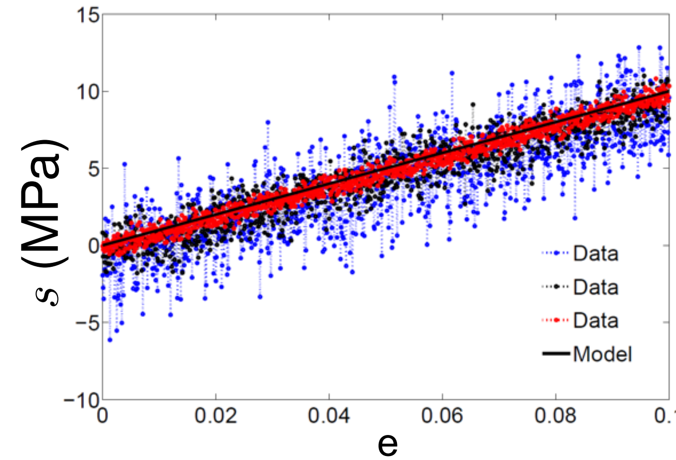
- Parameter Estimation: Parameters are considered to be random variables having associated densities.

Bayesian Inference: Simple Model

Example: Displacement-force relation (Hooke's Law)

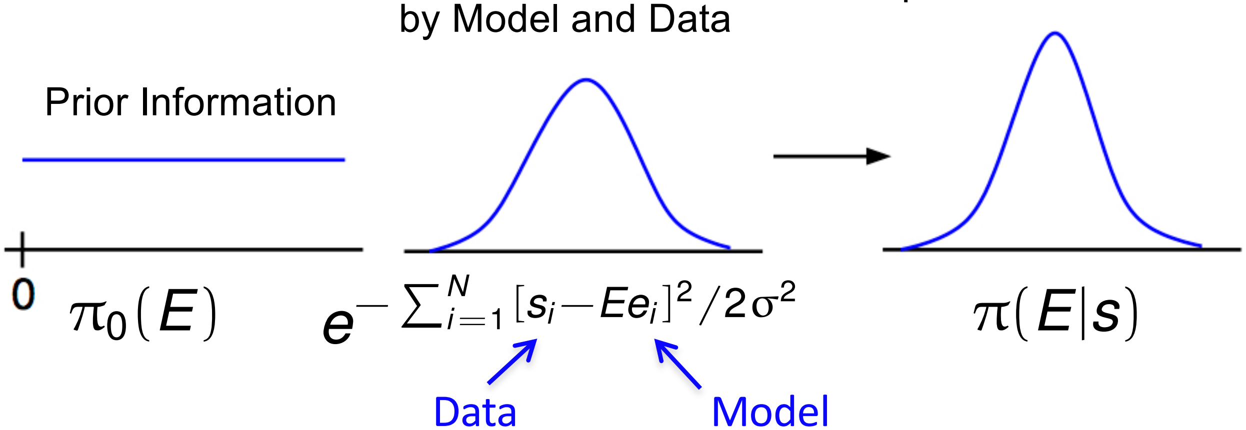
$$s_i = Ee_i + \varepsilon_i \;,\; i = 1, \ldots, N$$

$$\varepsilon_i \sim N(0, \sigma^2)$$



Parameter: Stiffness E

Strategy: Use model fit to data to update prior information



$$\pi_0(E) \qquad e^{-\sum_{i=1}^N [s_i - Ee_i]^2/2\sigma^2} \qquad \pi(E|s)$$

Non-normalized Bayes' Relation:

$$\pi(E|s) = e^{-\sum_{i=1}^N [s_i - Ee_i]^2/2\sigma^2} \pi_0(E)$$

Bayesian Inference

Bayes' Relation: Specifies posterior in terms of likelihood and prior

Likelihood: $e^{-\sum_{i=1}^{N}[s_i - Ee_i]^2/2\sigma^2}$, $q = E$

$\upsilon = [s_1, \ldots, s_N]$

Posterior Distribution

$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$

Prior Distribution

Normalization Constant

- Prior Distribution: Quantifies prior knowledge of parameter values
- Likelihood: Probability of observing a data given set of parameter values.
- Posterior Distribution: Conditional distribution of parameters given observed data.

Problem: Can require high-dimensional integration

- e.g., MFC Model: p = 20!
- Solution: Sampling-based Markov Chain Monte Carlo (MCMC) algorithms.
- Metropolis algorithms first used by nuclear physicists during Manhattan Project in 1940's to understand particle movement underlying first atomic bomb.

Bayesian Model Calibration

Bayesian Model Calibration:

- Parameters assumed to be random variables

$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$

Example: Coin Flip

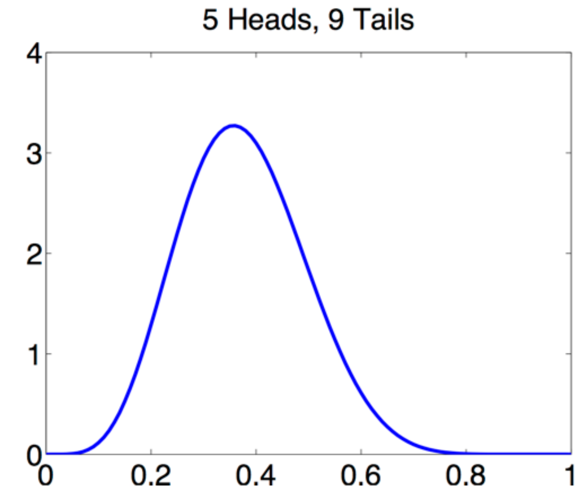
$$\Upsilon_i(\omega) = \begin{cases} 0 & , \quad \omega = T \\ 1 & , \quad \omega = H \end{cases}$$

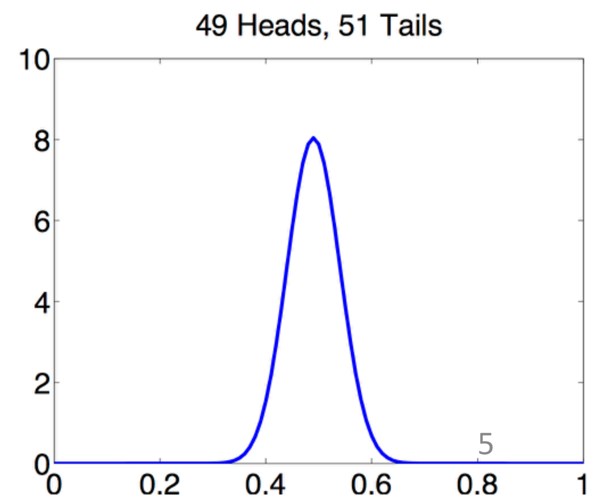
Likelihood:

$$\begin{aligned} \pi(\upsilon|q) &= \prod_{i=1}^{N} q^{\upsilon_i}(1-q)^{1-\upsilon_i} \\ &= q^{N_1}(1-q)^{N_0} \end{aligned}$$

Posterior with Noninformative Prior: $\pi_0(q) = 1$

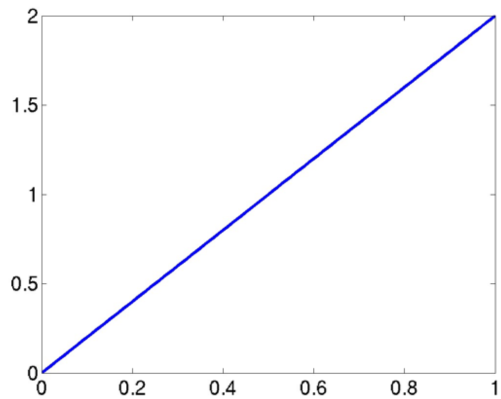
$$\pi(q|\upsilon) = \frac{q^{N_1}(1-q)^{N_0}}{\int_0^1 q^{N_1}(1-q)^{N_0}dq} = \frac{(N+1)!}{N_0!N_1!}q^{N_1}(1-q)^{N_0}$$



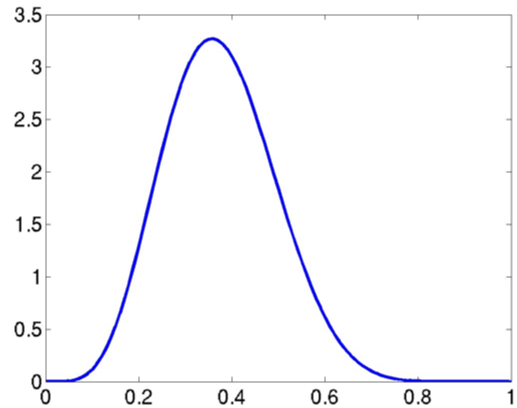


Bayesian Inference

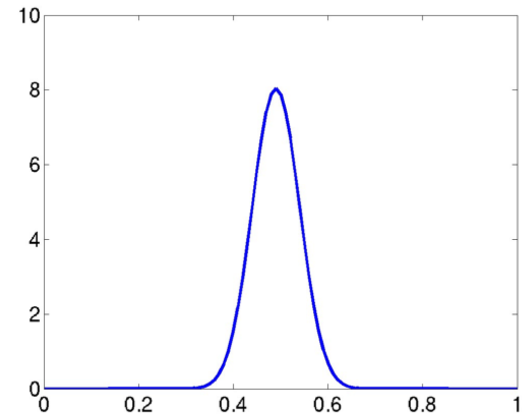
Example:



1 Head, 0 Tails



5 Heads, 9 Tails



49 Heads, 51 Tails

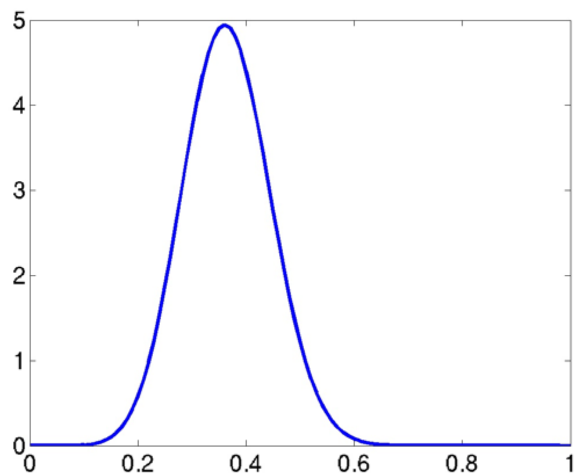
Note: For $N = 1$, frequentist theory would give probability 1 or 0

Bayesian Inference

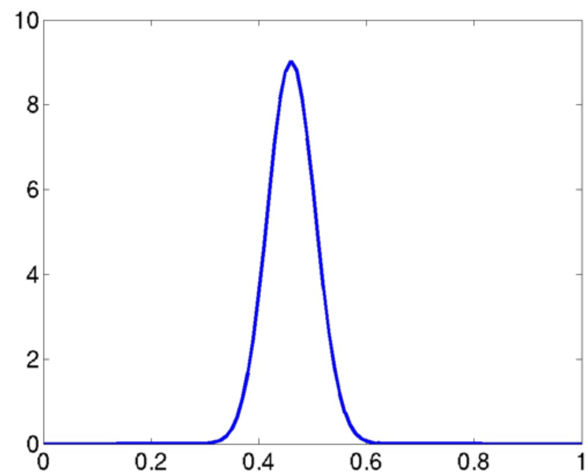
Example: Now consider

$$\pi_0(q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(q-\mu)^2/2\sigma^2}$$

with $\mu = .3$ and $\sigma = .1$.



5 Heads, 5 Tails



50 Heads, 50 Tails

Note: Poor informative prior incorrectly influences results for a long time.

Parameter Estimation Problem

Assumption: Assume that measurement errors are iid and $\varepsilon_i \sim N(0, \sigma^2)$

Likelihood:

$$\pi(\upsilon|q) = L(q, \sigma|\upsilon) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

where

$$SS_q = \sum_{i=1}^{n} [\upsilon_i - f_i(q)]^2$$

is the sum of squares error.

Parameter Estimation: Example

Example: Consider the spring model

$$\ddot{z} + C\dot{z} + Kz = 0$$

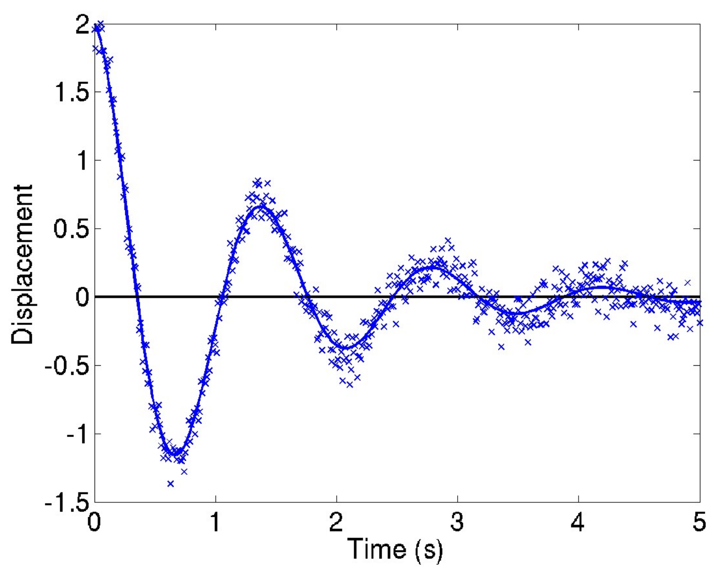
$$z(0) = 2\ ,\ \dot{z}(0) = -C$$

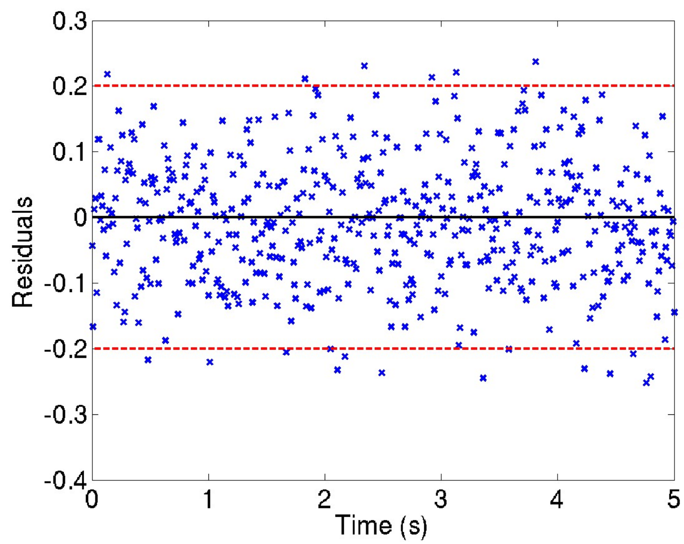
Note: Take $K = 20.5, C_0 = 1.5$

which has the solution

$$z(t) = 2e^{-Ct/2}\cos(\sqrt{K - C^2/4} \cdot t)$$

Take K to be known and $Q = C$. We also assume that $\varepsilon_i \sim N(0, \sigma_0^2)$ where $\sigma_0 = 0.1$.





Parameter Estimation: Example

Ordinary Least Squares: Here

$$\mathcal{X}(q) = \left[\frac{\partial y}{\partial C}(t_1, q), \cdots, \frac{\partial y}{\partial C}(t_n, q) \right]^T$$

where

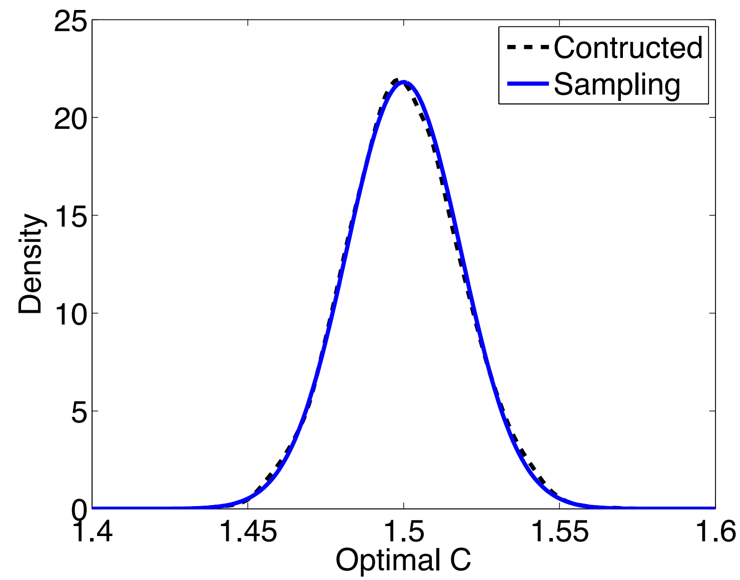
$$\frac{\partial y}{\partial C} = e^{-Ct/2} \left[\frac{Ct}{\sqrt{4K - C^2}} \sin\left(\sqrt{K - C^2/4} \cdot t\right) - t\cos\left(\sqrt{K - C^2/4} \cdot t\right) \right]$$

Then

$$V = \sigma_c^2 = \sigma_0^2 \left[\mathcal{X}^T(q) \mathcal{X}(q) \right]^{-1} = 3.35 \times 10^{-4}$$

so that

$$\widehat{C} \sim N\left(C_0, \sigma_c^2\right) \;,\; \sigma_c = 0.0183$$



Parameter Estimation: Example

Bayesian Inference: Employ the uniformed prior

$$\pi_0(q) = \chi_{[0,\infty)}(q)$$

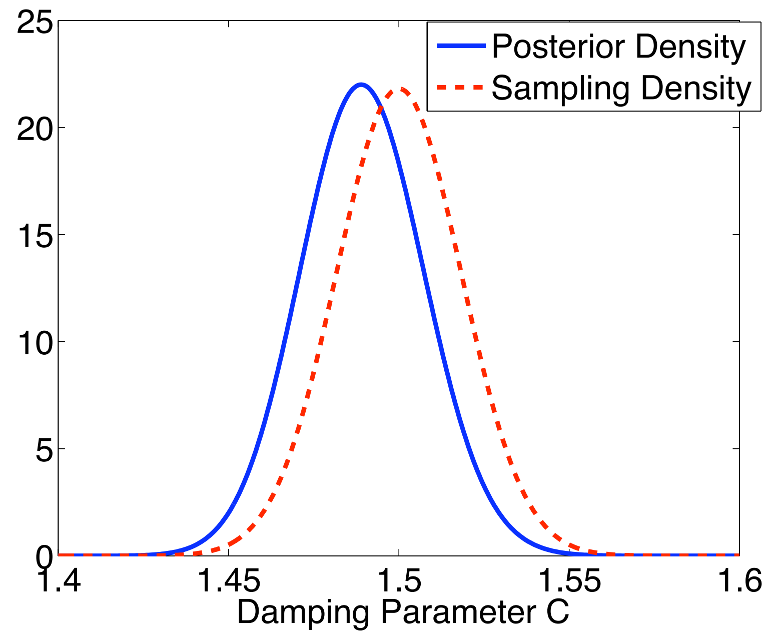
Posterior Distribution:

$$\pi(q|\upsilon) = \frac{e^{-SS_q/2\sigma_0^2}}{\int_0^\infty e^{-SS_\zeta/2\sigma_0^2} d\zeta} = \frac{1}{\int_0^\infty e^{-(SS_\zeta - SS_q)/2\sigma_0^2} d\zeta}$$

Issue: $e^{-SS_{q_{MAP}}} \approx 3 \times 10^{-113}$

Midpoint formula:

$$\pi(q|\upsilon) \approx \frac{1}{\sum_k e^{-(SS_{\zeta_i} - SS_q)w_i/2\sigma_0^2}}$$



Note:

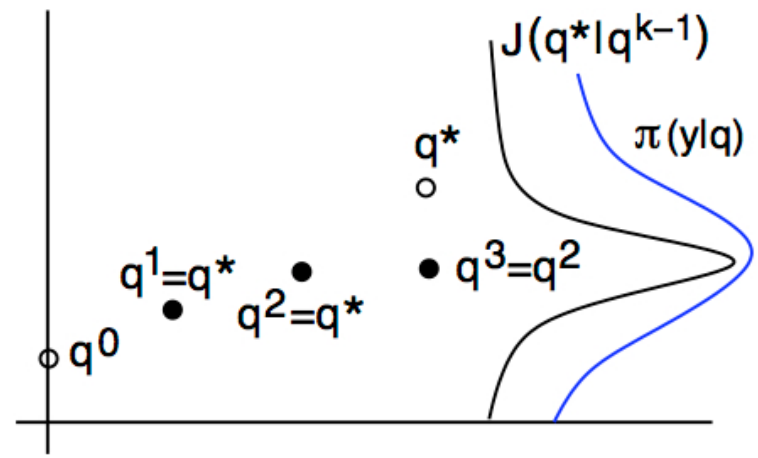
- Slow even for one parameter.
- Strategy: create Markov chain using random sampling so that created chain has the posterior distribution as its limiting (stationary) distribution.

Bayesian Model Calibration

Bayesian Model Calibration:

•Parameters considered to be random variables with associated densities.

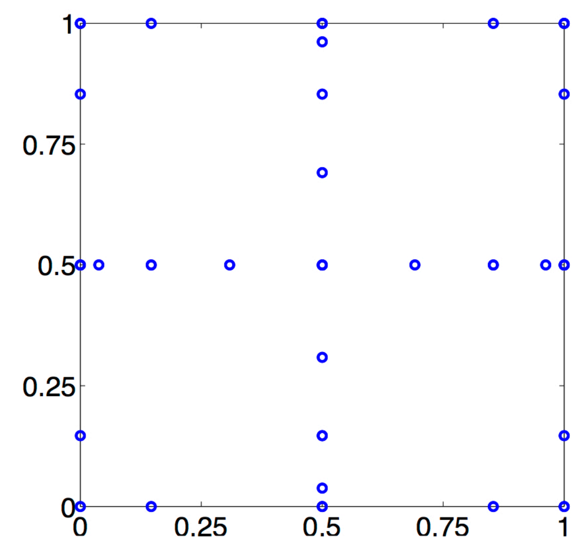
$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$



Problem:

•Often requires high dimensional integration;

- e.g., $p = 18$ for MFC model
- p = thousands to millions for some models



Strategies:

- Sampling methods
- Sparse grid quadrature techniques

Markov Chains

Definition: Sequence of random variables $X_1, X_2, \cdots$ that satisfy Markov property:
X_{n+1} depends only on X_n; that is

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

where x_i is the state of the chain at time i.

Note: A Markov chain is characterized by three components: a state space, an initial distribution, and a transition kernel.

State Space: Range of X_i: Set of all possible values

Initial Distribution: (Mass)

$$p^0 = [p_1^0, p_2^0, \cdots, p_n^0] \quad , \quad p_i^0 = P(X_0 = x_i)$$

Transition Probability: (Markov Kernel)

$$p_{ij} = P(X_{n+1} = x_j | X_n = x_i)$$

$$p_{ij}^{(n)} = P(X_{m+n} = x_j | X_m = x_i) \quad \text{(n-step transition probability)}$$

$$P = [p_{ij}] \quad , \quad P_n = [p_{ij}^{(n)}]$$

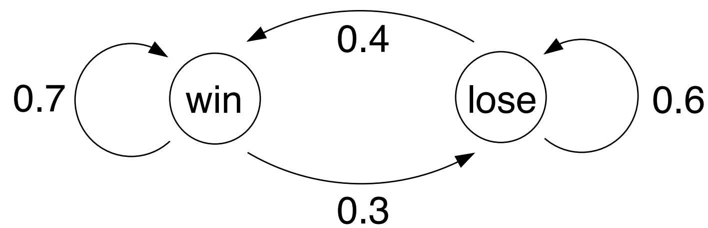
Markov Chain Techniques

Markov Chain: Sequence of events where current state depends only on last value.

Baseball: States are S = $\{$win,lose$\}$. Initial state is $p^0 = [0.8, 0.2]$.

• Assume that team which won last game has 70% chance of winning next game and 30% chance of losing next game.

• Assume losing team wins 40% and loses 60% of next games.



• Percentage of teams who win/lose next game given by

$$p^1 = [0.8\ ,\ 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [0.64\ ,\ 0.36]$$

• Question: does the following limit exist?

$$p^n = [0.8\ ,\ 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^n$$

Markov Chain Techniques

Baseball Example: Solve constrained relation

$$\pi = \pi P \quad , \quad \sum \pi_i = 1$$

$$\Rightarrow [\pi_{win} \, , \, \pi_{lose}] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [\pi_{win} \, , \, \pi_{lose}] \quad , \quad \pi_{win} + \pi_{lose} = 1$$

to obtain

$$\pi = [0.5714 \, , \, 0.4286]$$

Markov Chain Techniques

Baseball Example: Solve constrained relation

$$\pi = \pi P \quad , \quad \sum \pi_i = 1$$

$$\Rightarrow [\pi_{win} \, , \, \pi_{lose}] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [\pi_{win} \, , \, \pi_{lose}] \quad , \quad \pi_{win} + \pi_{lose} = 1$$

to obtain

$$\pi = [0.5714 \, , \, 0.4286]$$

Alternative: Iterate to compute solution

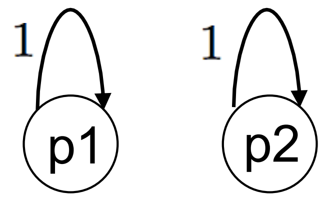
n	p^n	n	p^n	n	p^n
0	$[0.8000 \, , \, 0.2000]$	4	$[0.5733 \, , \, 0.4267]$	8	$[0.5714 \, , \, 0.4286]$
1	$[0.6400 \, , \, 0.3600]$	5	$[0.5720 \, , \, 0.4280]$	9	$[0.5714 \, , \, 0.4286]$
2	$[0.5920 \, , \, 0.4080]$	6	$[0.5716 \, , \, 0.4284]$	10	$[0.5714 \, , \, 0.4286]$
3	$[0.5776 \, , \, 0.4224]$	7	$[0.5715 \, , \, 0.4285]$		

Notes:

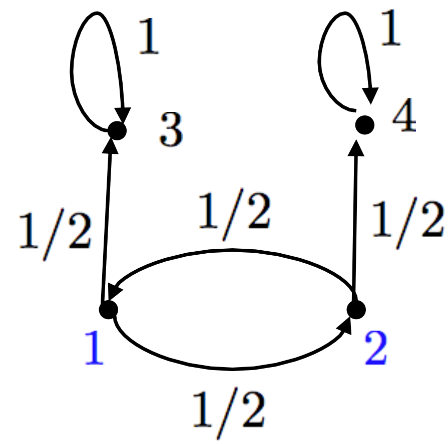
- Forms basis for Markov Chain Monte Carlo (MCMC) techniques
- Goal: construct chains whose stationary distribution is the posterior density

Irreducible Markov Chains

Reducible Markov Chain:



$$p^0 = [p1\ ,\ p2] = \pi$$



$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

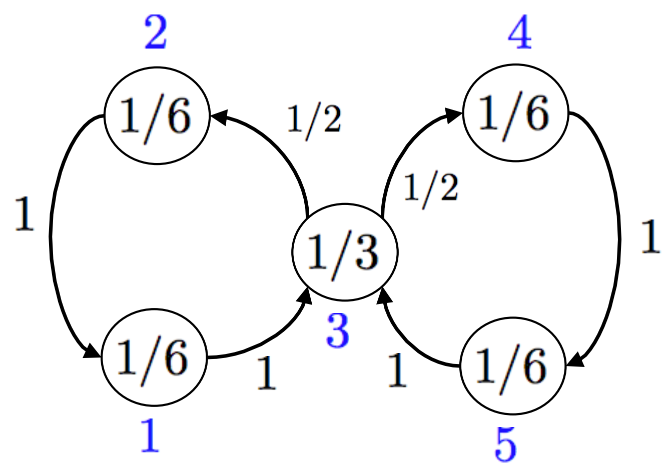
Note: Limiting distribution not unique if chain is reducible.

Irreducible: A Markov chain is *irreducible* if any state x_j and be reached from any state x_i in a finite number of steps; that is

$p_{ij}^{(n)} > 0$ for all states in finite n

Periodic Markov Chains

Example:



$$P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\pi = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

Note: Chain returns to state 1 at steps $3, 6, 9, \cdots$ so Period = 3

Note: Probability mass "cycles" through chain so no convergence

Periodicity: A Markov chain is *periodic* if parts of the state space are visited at regular intervals. The period *k* is defined as

$$\begin{aligned} k &= \gcd\left\{n | p_{ii}^{(n)} > 0\right\} \\ &= \gcd\left\{n | P(X_{m+n} = x_i | X_m = x_i) > 0\right\} \end{aligned}$$

- The chain is aperiodic if $k = 1$.

Stationary Distribution

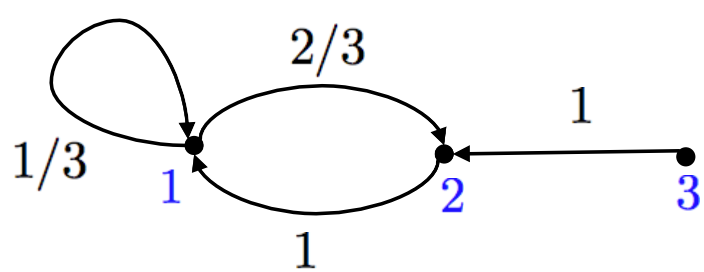
Theorem: A finite, homogeneous Markov chain that is irreducible and aperiodic has a unique stationary distribution π and the chain will converge in the sense of distributions from any initial distribution p^0.

Recurrence (Persistence): A state x_i is recurrent (persistent) if the probability of returning to x_i is 1; that is,

$$P(X_{m+n} = x_i \text{ for some } n \geq 1 | X_m = x_i) = 1$$

- It is *transient* if probability strictly less than 1

Example: State 3 is transient



Ergodicity: A state is termed *ergodic* if it is aperiodic and recurrent. If all states of an irreducible Markov chain are ergodic, the chain is said to be *ergodic*.

Matrix Theory

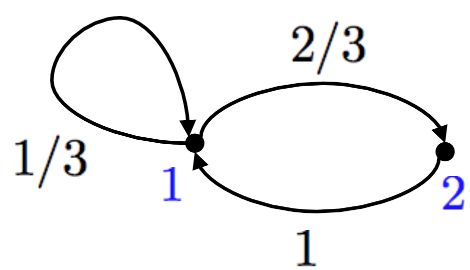
Definition: A matrix $A \in \mathbb{R}^{(n \times n)}$ is

(i) Nonnegative, denoted $A \geq 0$, if $a_{ij} \geq 0$ for all i, j

(ii) Strictly positive, denoted $A > 0$, if $a_{ij} > 0$ for all i, j

Lemma: Let P be the transition matrix of an ergodic finite Markov chain with state space S. Then for some $N_0 \geq 1$, $P_n > 0$ for all $n > N_0$.

Example:



$$P = \begin{bmatrix} 1/3 & 2/3 \\ 1 & 0 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 7/9 & 2/9 \\ 1/3 & 2/3 \end{bmatrix}$$

Matrix Theory

Theorem (Perron-Frobenius): For any strictly positive matrix $A > 0$, there exist $\lambda_0 > 0$ and $x_0 > 0$ such that

(i) $Ax_0 = \lambda_0 x_0$

(ii) If $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $|\lambda| < \lambda_0$

(iii) λ_0 has geometric and algebraic multiplicity 1

Corollary 1: If $A \geq 0$ is a nonnegative matrix such that $A^n > 0$, then theorem also applies to A.

Proposition: Let $A > 0$ be a strictly positive $n \times n$ matrix with row and column sums

$$r_i = \sum_j a_{ij} \quad , \quad c_j = \sum_i a_{ij} \; , \; i, j = 1, \cdots, n$$

Then

$$\min_i r_i \leq \lambda_0 \leq \max_i r_i \quad , \quad \min_j c_j \leq \lambda_0 \leq \max_j c_j$$

Stationary Distribution

Corollary: Let $P \geq 0$ be the transition matrix of an ergodic Markov chain. Then there exists a unique stationary distribution π such that $\pi P = \pi$.

Proof: By Lemma and Corollary 1, P has a largest eigenvalue $\lambda_0 = 1$.
Since multiplicity is 1, unique π such that $\pi P = \pi$ and $\sum_i \pi_i = 1$.

Convergence: Express

$$UPV = \Lambda = \begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

where $1 > |\lambda_2| \geq \cdots \geq |\lambda_n|, V = U^{-1}$

Note:

$$P^n = V \begin{bmatrix} 1 & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_n^n \end{bmatrix} U \rightarrow V \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} U$$

Stationary Distribution

Note:

$$UP = \Lambda U \Rightarrow \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \vdots & & \vdots \\ & & \end{bmatrix} \begin{bmatrix} & & \\ & P & \\ & & \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \vdots & & \vdots \\ & & \end{bmatrix}$$

and

$$V = U^{-1} = \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \vdots & & \vdots \\ & & \end{bmatrix} \begin{bmatrix} 1 & \cdots & \\ \vdots & & \\ 1 & \cdots & \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

Thus

$$\begin{aligned} \lim_{n\to\infty} p^n &= \lim_{n\to\infty} p^0 P^n \\ &= \lim_{n\to\infty} \begin{bmatrix} p_1^0 & \cdots & p_n^0 \end{bmatrix} \begin{bmatrix} 1 & \cdots & \\ \vdots & & \\ 1 & \cdots & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \vdots & & \vdots \\ & & \end{bmatrix} \\ &= \begin{bmatrix} p_1^0 & \cdots & p_n^0 \end{bmatrix} \begin{bmatrix} 1 & \cdots & \\ \vdots & & \\ 1 & \cdots & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \vdots & & \vdots \\ & & \end{bmatrix} \\ &= \begin{bmatrix} \pi_1 \cdots \pi_n \end{bmatrix} \\ &= \pi \end{aligned}$$

Detailed Balance Conditions

Situation: We can prove convergence of π such that $\pi P = \pi$. However, it doesn't give us an algorithm to construct it. This is provided by *detailed balance conditions*.

Reversible Chains: A Markov chain determined by the transition matrix $P = [p_{ij}]$ is reversible if there is a distribution π that satisfies the detailed balance conditions

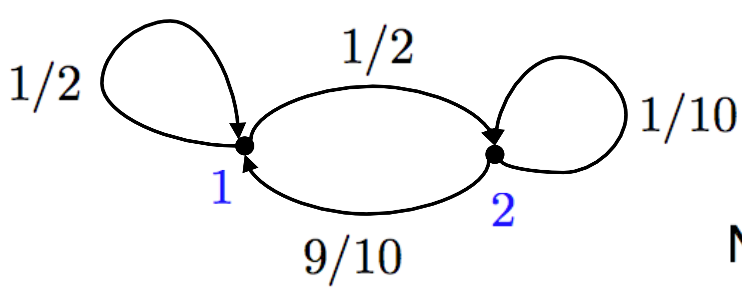
$$\pi_i p_{ij} = \pi_j p_{ji}$$

Note: Detailed balance implies that

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

$$\Rightarrow \pi P = \pi$$

Example:



$$P = \begin{bmatrix} 1/2 & 1/2 \\ 9/10 & 1/10 \end{bmatrix}$$

$$\pi = \begin{bmatrix} 9/14 & 5/14 \end{bmatrix}$$

Note: $\frac{1}{2} \cdot \frac{9}{14} = \frac{9}{10} \cdot \frac{5}{14}$ so detailed balance satisfied

Markov Chain Monte Carlo Methods

Strategy: Markov chain simulation used when it is impossible, or computationally prohibitive, to sample q directly from

$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$

- Create a Markov process whose stationary distribution is $\pi(q|\upsilon)$.

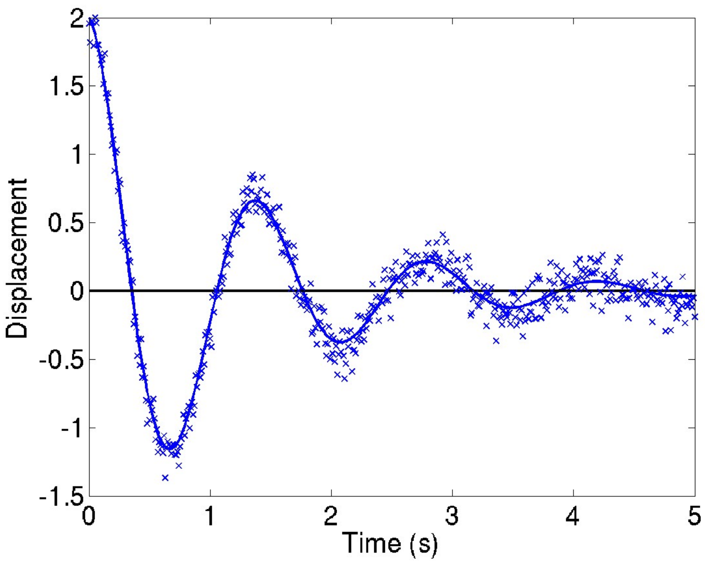
Note:

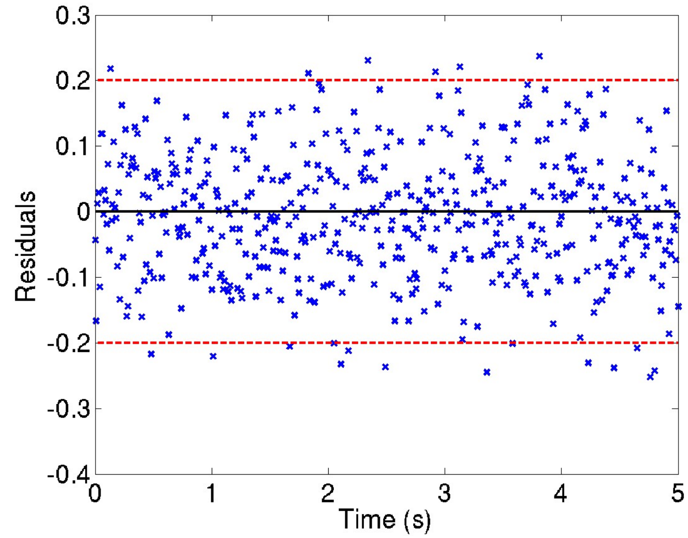
• In Markov chain theory, we are given a Markov chain, P, and we construct its equilibrium distribution.

• In MCMC theory, we are "given" a distribution and we want to construct a Markov chain that is reversible with respect to it.

Model Calibration Problem

Assumption: Assume that measurement errors are iid and $\varepsilon_i \sim N(0, \sigma^2)$





Likelihood:

$$\pi(\upsilon|q) = L(q, \sigma|\upsilon) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

where

$$SS_q = \sum_{i=1}^{n} [\upsilon_i - f_i(q)]^2$$

is the sum of squares error.

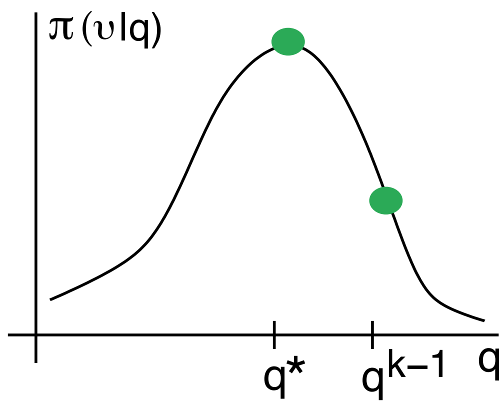
Markov Chain Monte Carlo Methods

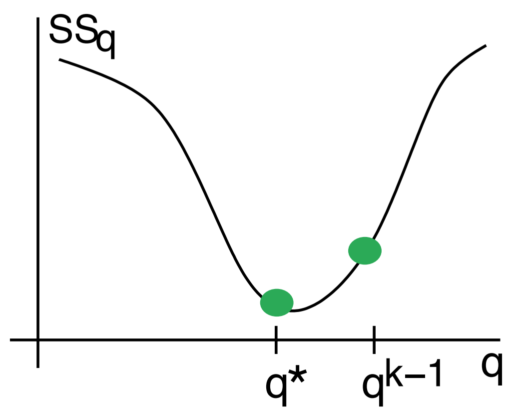
Strategy:

- Sample values from proposal distribution $J(q^*|q^{k-1})$ that reflects geometry of posterior distribution
- Compute $r(q^*|q^{k-1}) = \frac{\pi(\upsilon|q^*)\pi_0(q^*)}{\pi(\upsilon|q^{k-1})\pi_0(q^{k-1})}$
  - If $r \geqslant 1$, accept with probability $\alpha = 1$
  - If $r < 1$, accept with probability $\alpha = r$

Intuition: Consider flat prior $\pi_0(q) = 1$ and Gaussian observation model

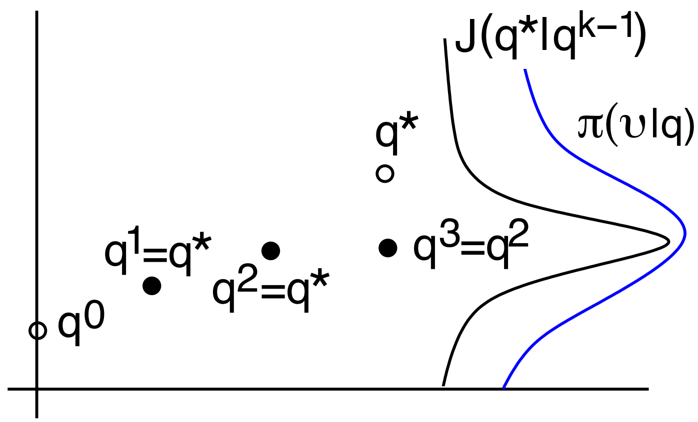
$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2} \quad SS_q = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$$

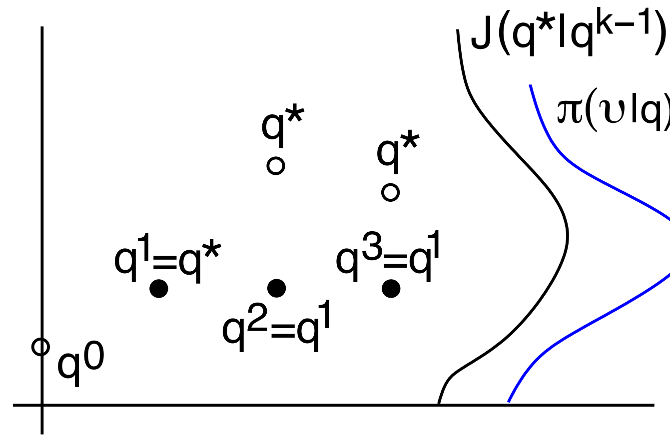


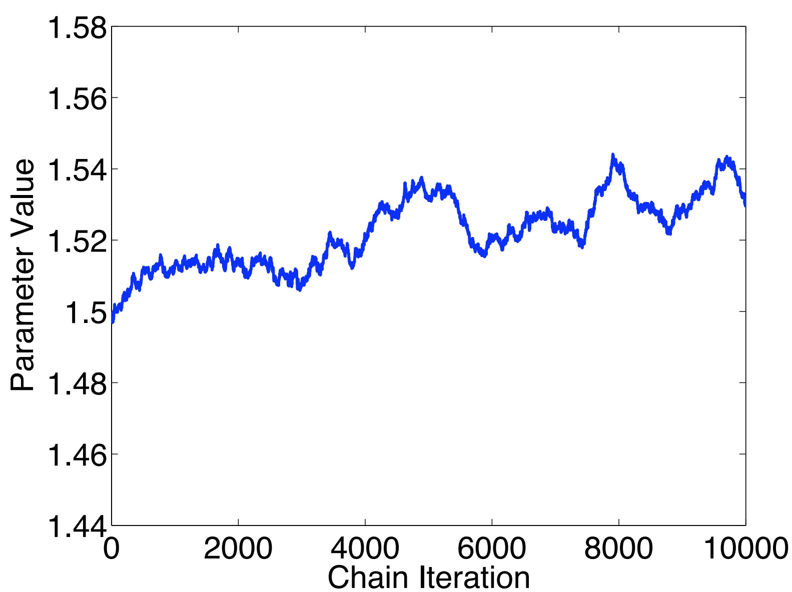


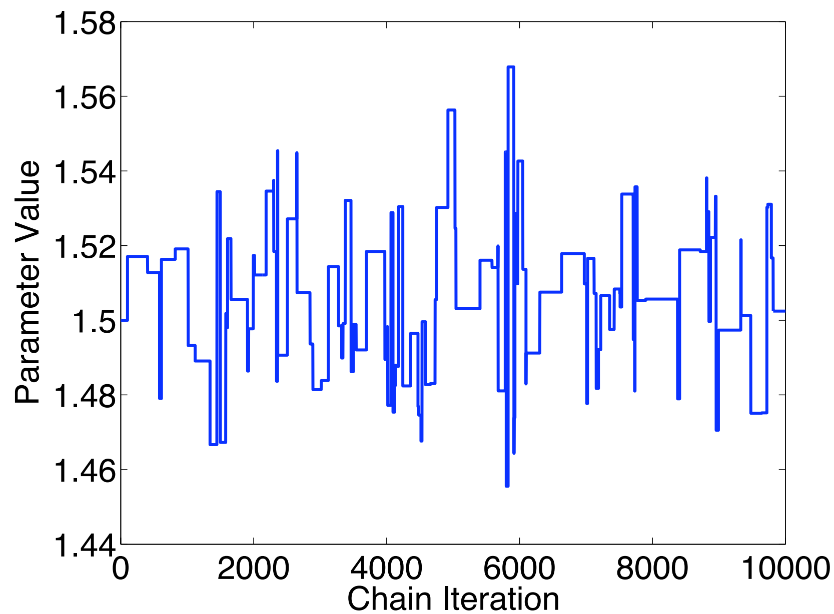
Markov Chain Monte Carlo Methods

Note: Narrower proposal distribution yields higher probability of acceptance.





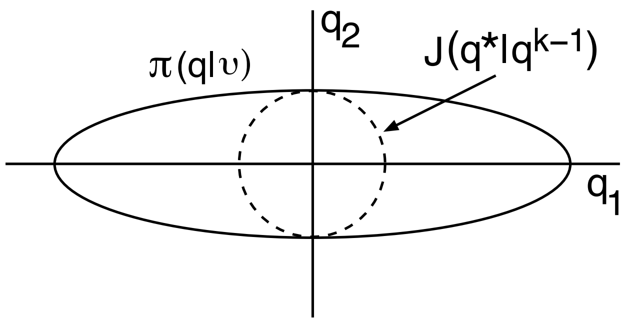




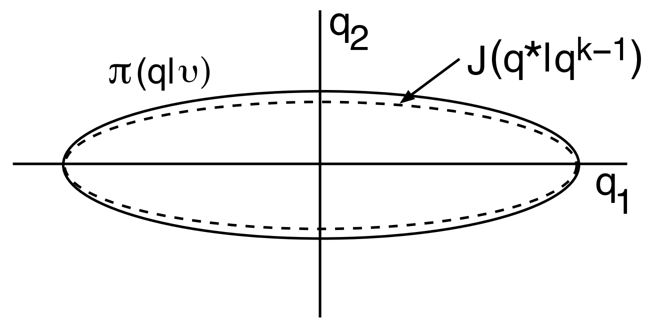
Proposal Distribution

Proposal Distribution: Significantly affects mixing

- Too wide: Too many points rejected and chain stays still for long periods;
- Too narrow: Acceptance ratio is high but algorithm is slow to explore parameter space
- Ideally, it should have similar "shape" to posterior distribution.



(a)



(b)

Problem:

- Anisotropic posterior, isotropic proposal;
- Efficiency nonuniform for different parameters

Result:

- Recovers efficiency of univariate case

Proposal Distribution

Proposal Distribution: Two basic approaches

- Choose a fixed proposal function
  - o Independent Metropolis
- Random walk (local Metropolis)

$$q^* = q^{k-1} + Rz$$

o Two (of several) choices: $z \sim N(0,1)$

$$\text{(i)}\ R = cI \Rightarrow q^* \sim N(q^{k-1}, cI)$$

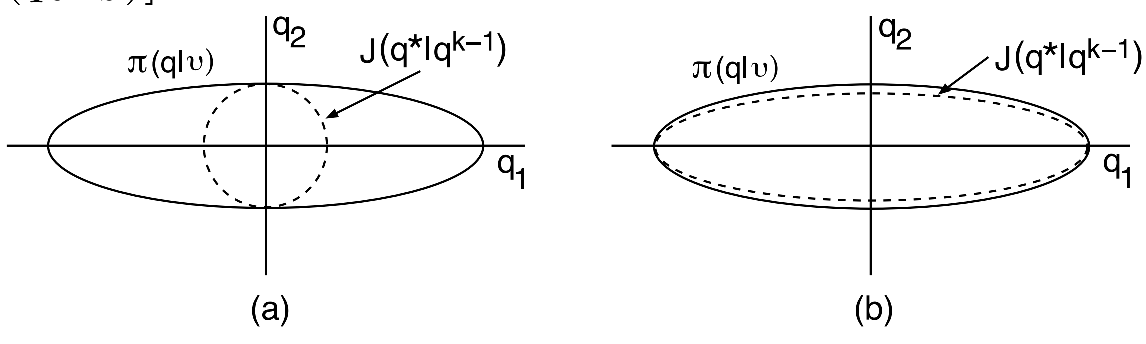
$$\text{(ii)}\ R = \text{chol}(V) \Rightarrow q^* \sim N(q^{k-1}, V)$$

where

Sensitivity Matrix

$$V = \sigma^2_{OLS}\left[\mathcal{X}^T(q_{OLS})\mathcal{X}(q_{OLS})\right]^{-1}$$

$$\sigma^2_{OLS} = \frac{1}{n-p}\sum_{i=1}^n [\upsilon_i - f_i(q_{OLS})]^2$$



Metropolis Algorithm

Metropolis Algorithm: [Metropolis and Ulam, 1949]

1. Initialization: Choose an initial parameter value q^0 that satisfies $\pi(q^0|\upsilon) > 0$.
2. For $k = 1, \cdots, M$

   (a) For $z \sim N(0,1)$, construct the candidate

   $$q^* = q^{k-1} + Rz$$

   where R is the Cholesky decomposition of V or D. This ensures that

   $$q^* \sim N(q^{k-1}, V) \;\text{ or }\; q^* \sim N(q^{k-1}, D).$$

   (b) Compute the ratio

   $$r(q^*|q^{k-1}) = \frac{\pi(q^*|\upsilon)}{\pi(q^{k-1}|\upsilon)} = \frac{\pi(\upsilon|q^*)\pi_0(q^*)}{\pi(\upsilon|q^{k-1})\pi_0(q^{k-1})}. \tag{1}$$

   (c) Set

   $$q^k = \begin{cases} q^* & , \quad \text{with probability } \alpha = \min(1, r) \\ q^{k-1} & , \quad \text{else.} \end{cases}$$

   That is, we accept q^* with probability 1 if $r \geq 1$ and we accept it with probability r if $r < 1$.

Sampling Error Variance

Strategy: Treat error variance σ^2 as parameter to be sampled.

Definition: The property that the prior and posterior distributions have the same parametric form is termed *conjugacy.*

Note: The likelihood

$$\pi(\upsilon, q|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

has the conjugate prior

$$\pi_0(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{\beta/\sigma^2}$$

The posterior is

$$\pi(\sigma^2|q, \upsilon) \propto (\sigma^2)^{-(\alpha+1+n/2)} e^{-(\beta+SS_q/2)/\sigma^2}$$

so that

$$\sigma^2|(\upsilon, q) \sim \text{Inv-gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{SS_q}{2}\right)$$

or

$$\sigma^2|(\upsilon, q) \sim \text{Inv-gamma}\left(\frac{n_s + n}{2}, \frac{n_s\sigma_s^2 + SS_q}{2}\right)$$

Note:

- n_0 taken small;

  e.g., $n_0 = 1$ or $n_0 = .01$
- Take $\sigma_s^2 = s_{k-1}^2 = \frac{R_{k-1}^T R_{k-1}}{n-p}$

Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R

1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$

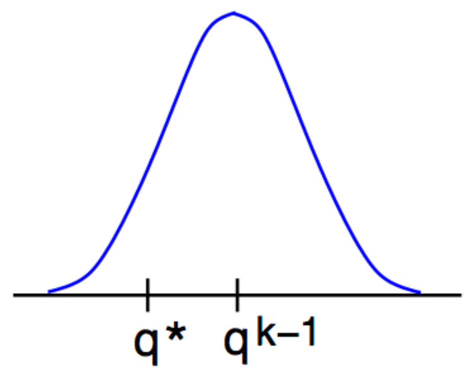
Example: Helmholtz energy

$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \longleftarrow \varepsilon_i \sim N(0, \sigma^2)$$

$$= \alpha_1 P_i^2 + \alpha_{11} P_i^4 + \varepsilon_i$$

Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R

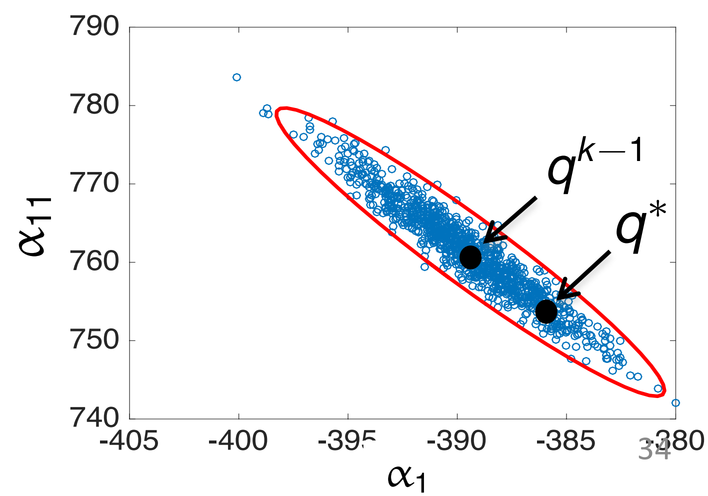


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, \underline{V})$

Example: Helmholtz energy

$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \longleftarrow \varepsilon_i \sim N(0, \sigma^2)$$
$$= \alpha_1 P_i^2 + \alpha_{11} P_i^4 + \varepsilon_i$$

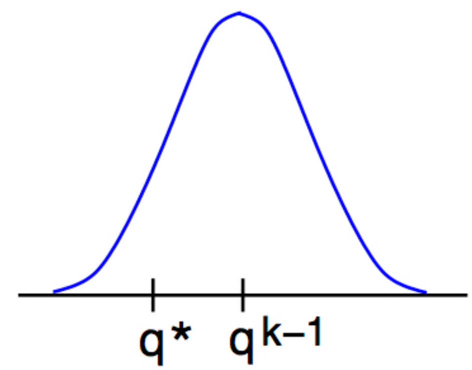
Recall: Covariance V incorporates geometry



Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R

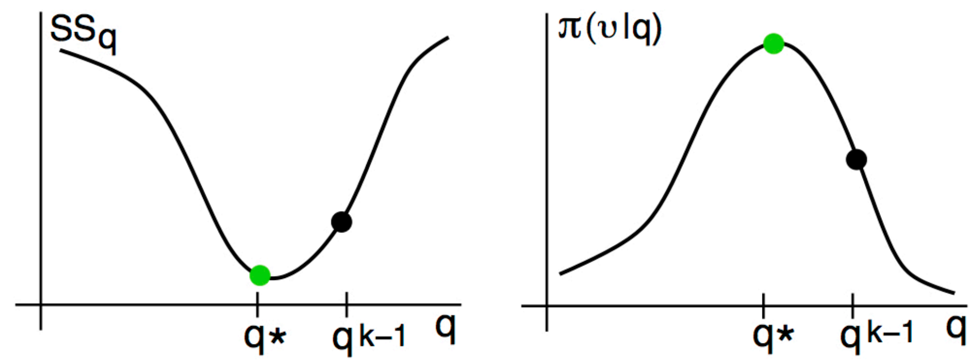
1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$



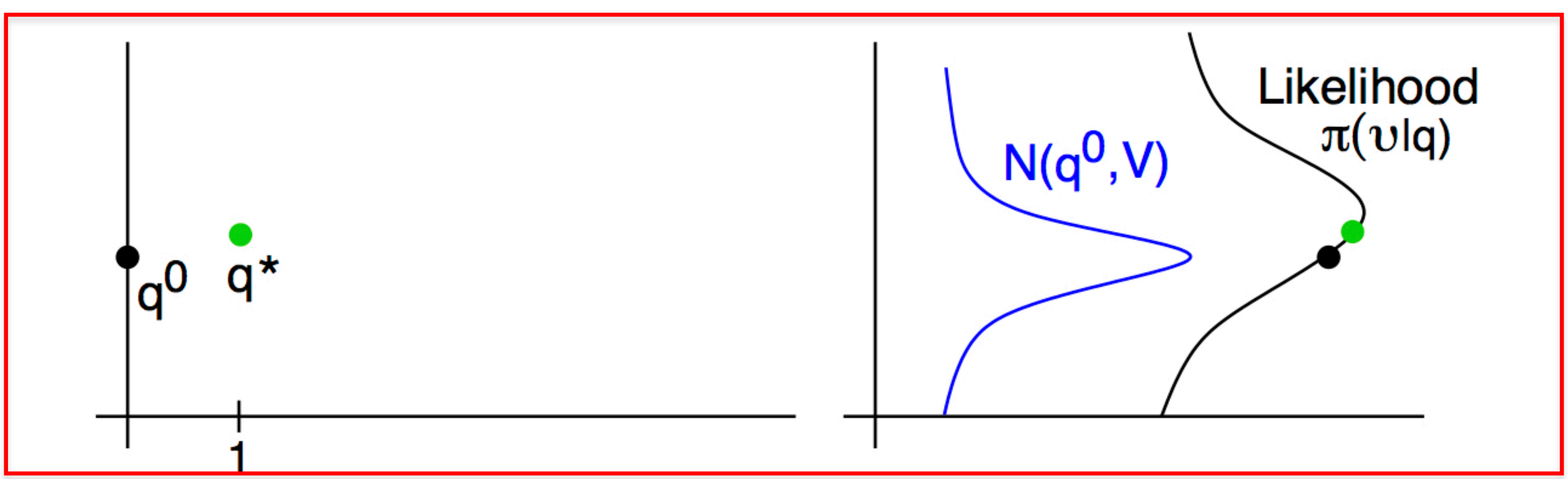
   (b) Compute likelihood

$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

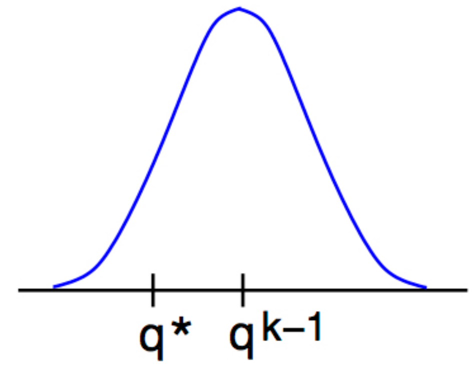


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

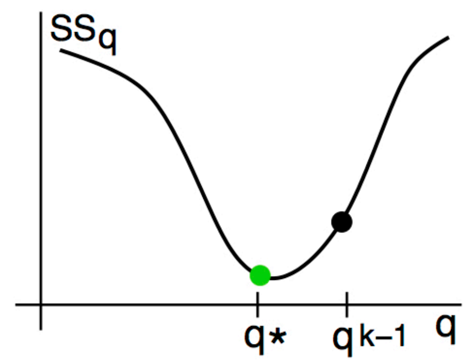
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

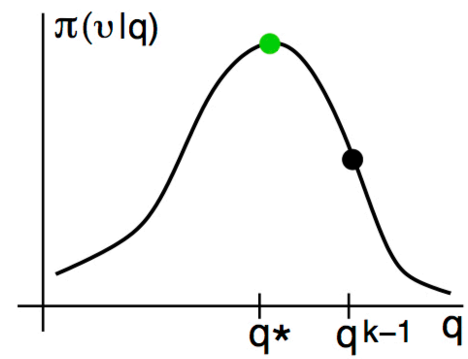


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

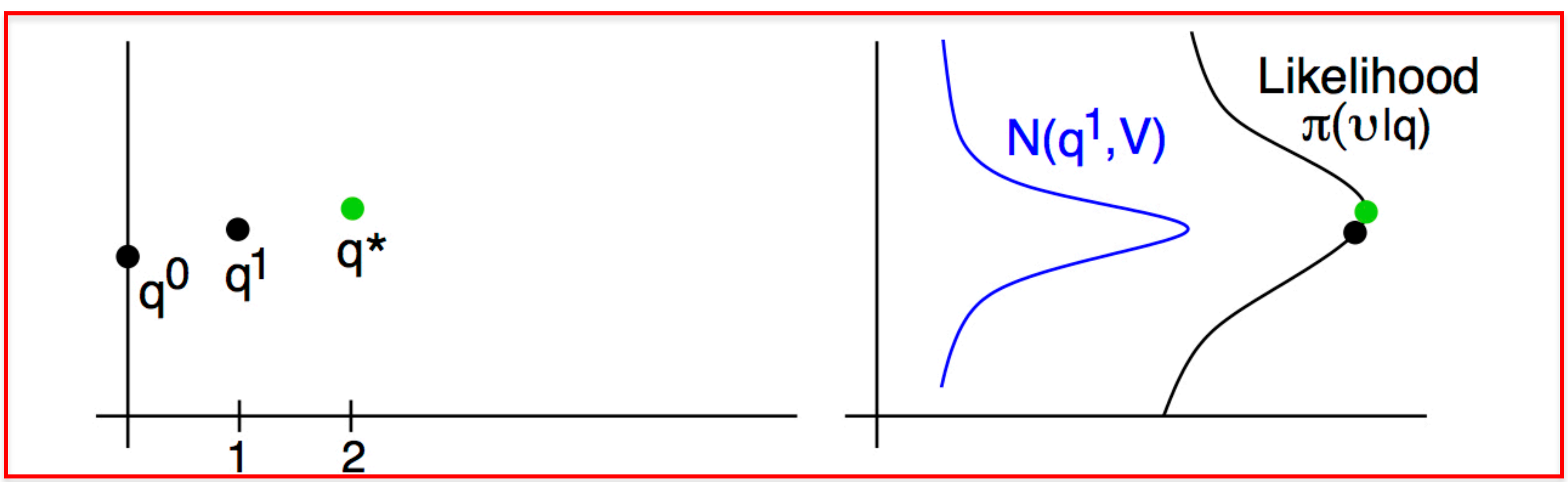
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$



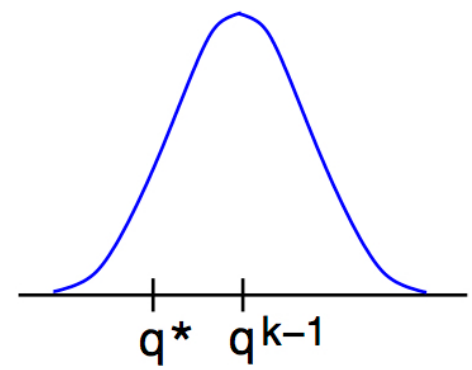


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

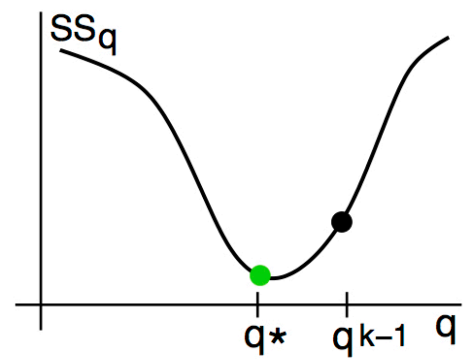
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

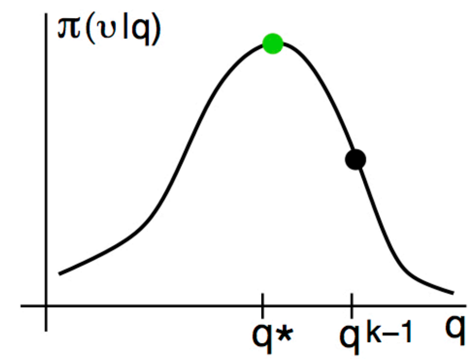


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

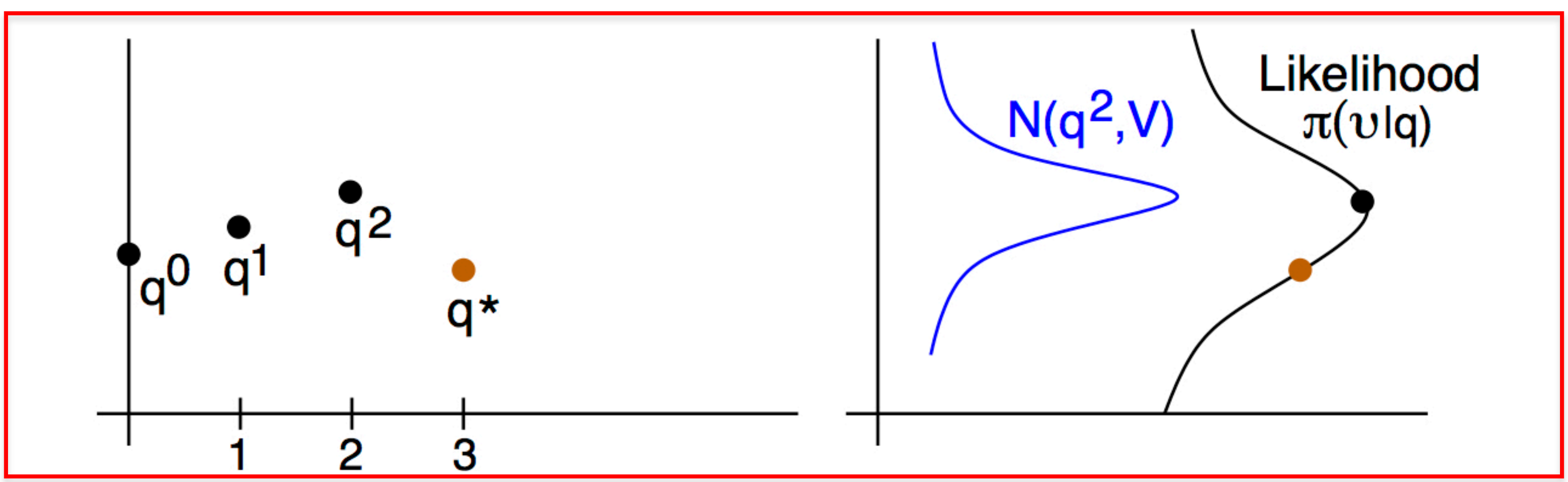
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$





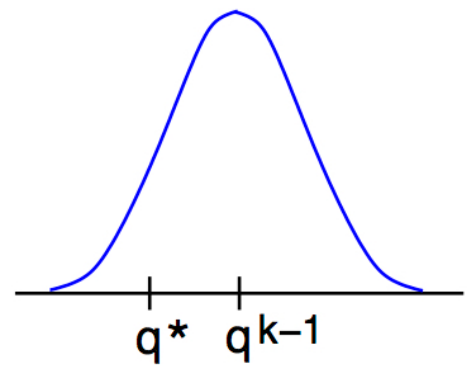
   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R

1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$



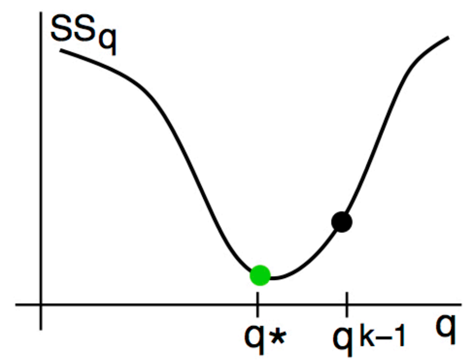
2. For $k = 1, \ldots, M$

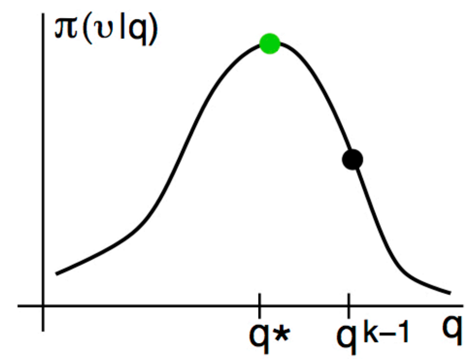
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$

   (b) Compute likelihood

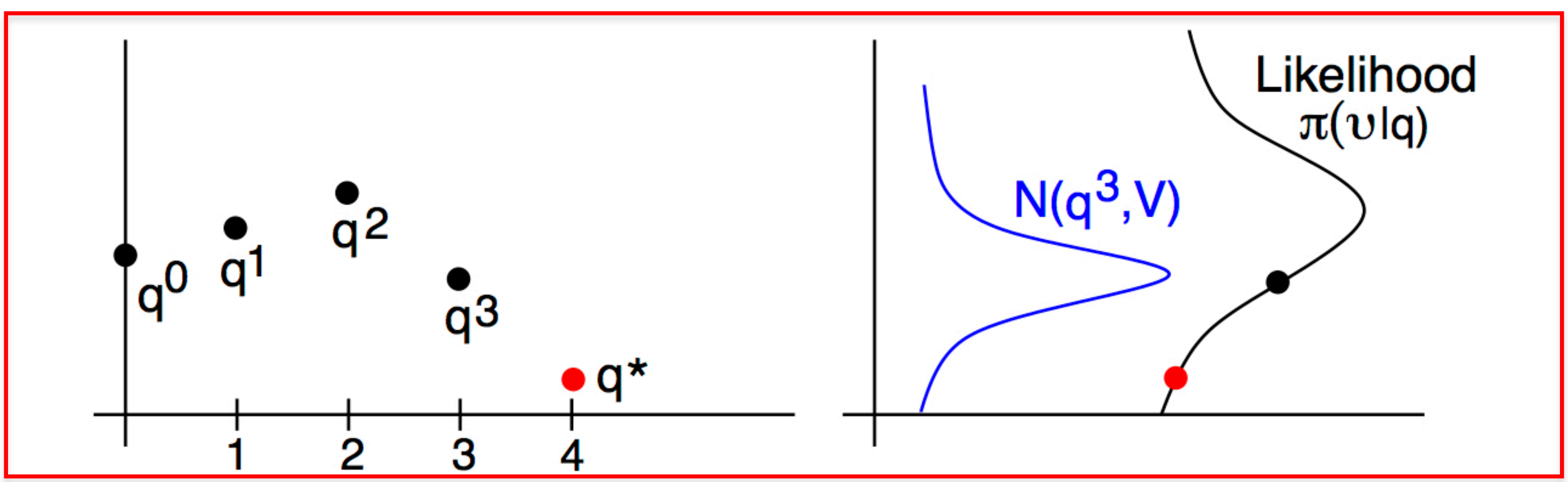
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$



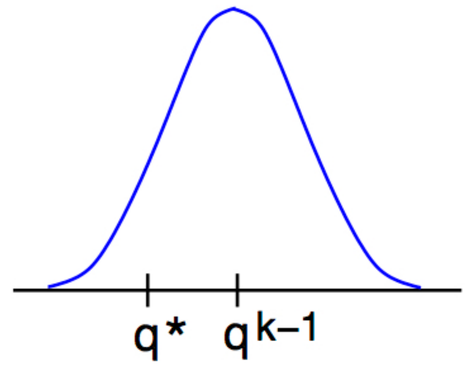


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R



1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$

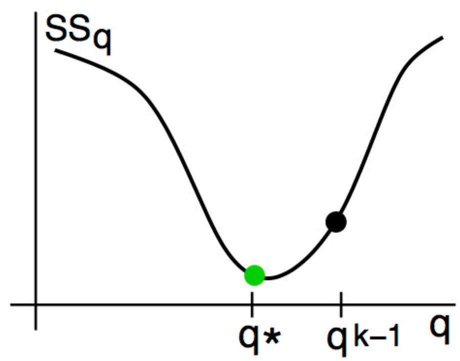
2. For $k = 1, \ldots, M$

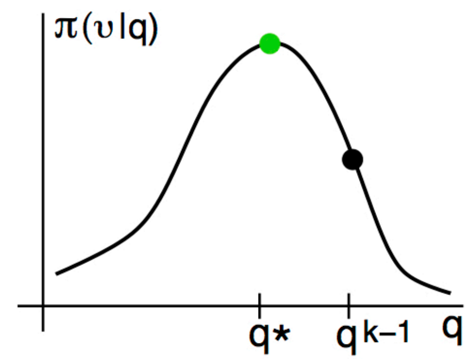
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$

   (b) Compute likelihood

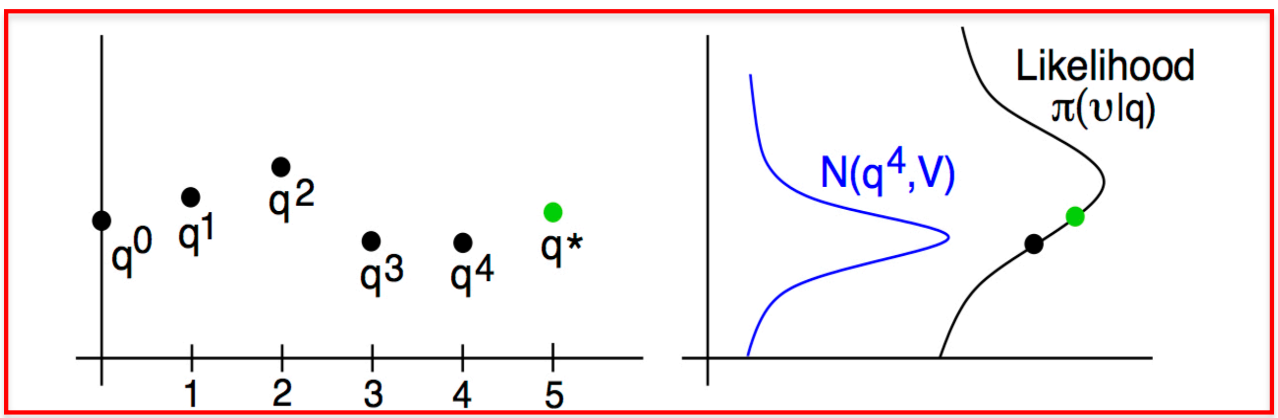
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$





   (c) Accept q^* with probability dictated by likelihood



Note:

- Delayed Rejection: Shrink proposal: γV
- Adaptive Metropolis: Update proposal as samples are accepted

Random Walk Metropolis

Example: We revisit the spring model

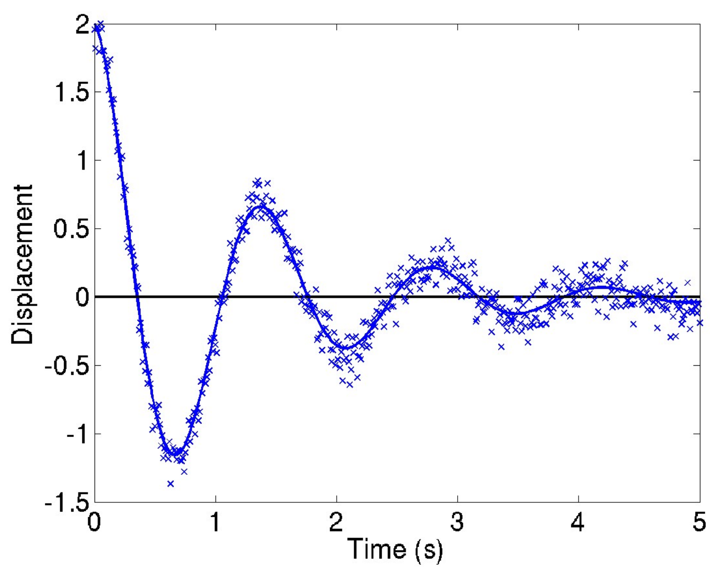
$$\ddot{z} + C\dot{z} + Kz = 0$$

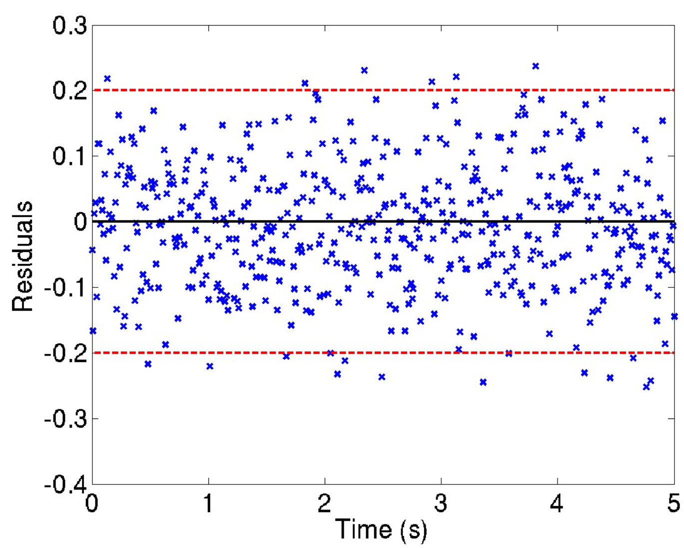
$$z(0) = 2\ ,\ \ \dot{z}(0) = -C$$

which has the solution

$$z(t) = 2e^{-Ct/2}\cos(\sqrt{K - C^2/4} \cdot t)$$

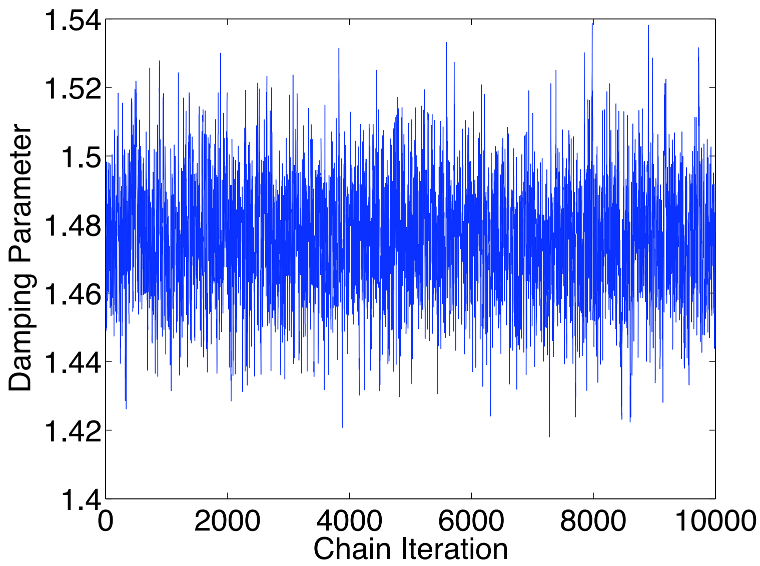
We assume that $\varepsilon_i \sim N(0, \sigma_0^2)$ where $\sigma_0 = 0.1$.

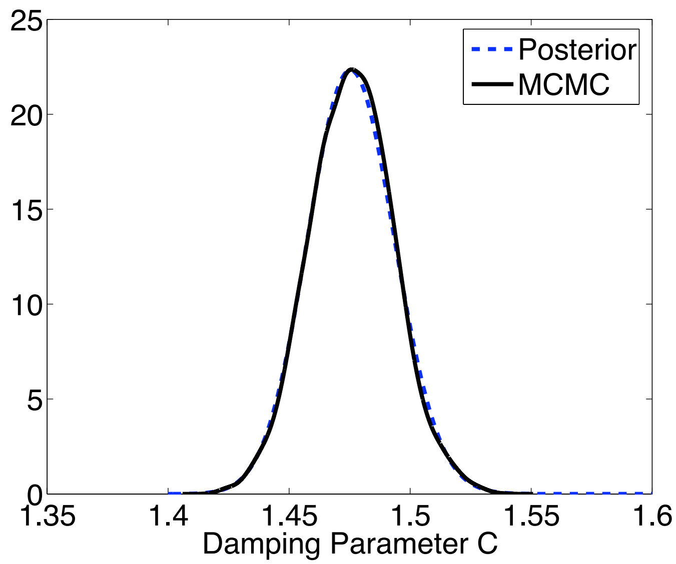




Random Walk Metropolis

Case i: Take K =20.5 and $Q = [C, \sigma^2]$



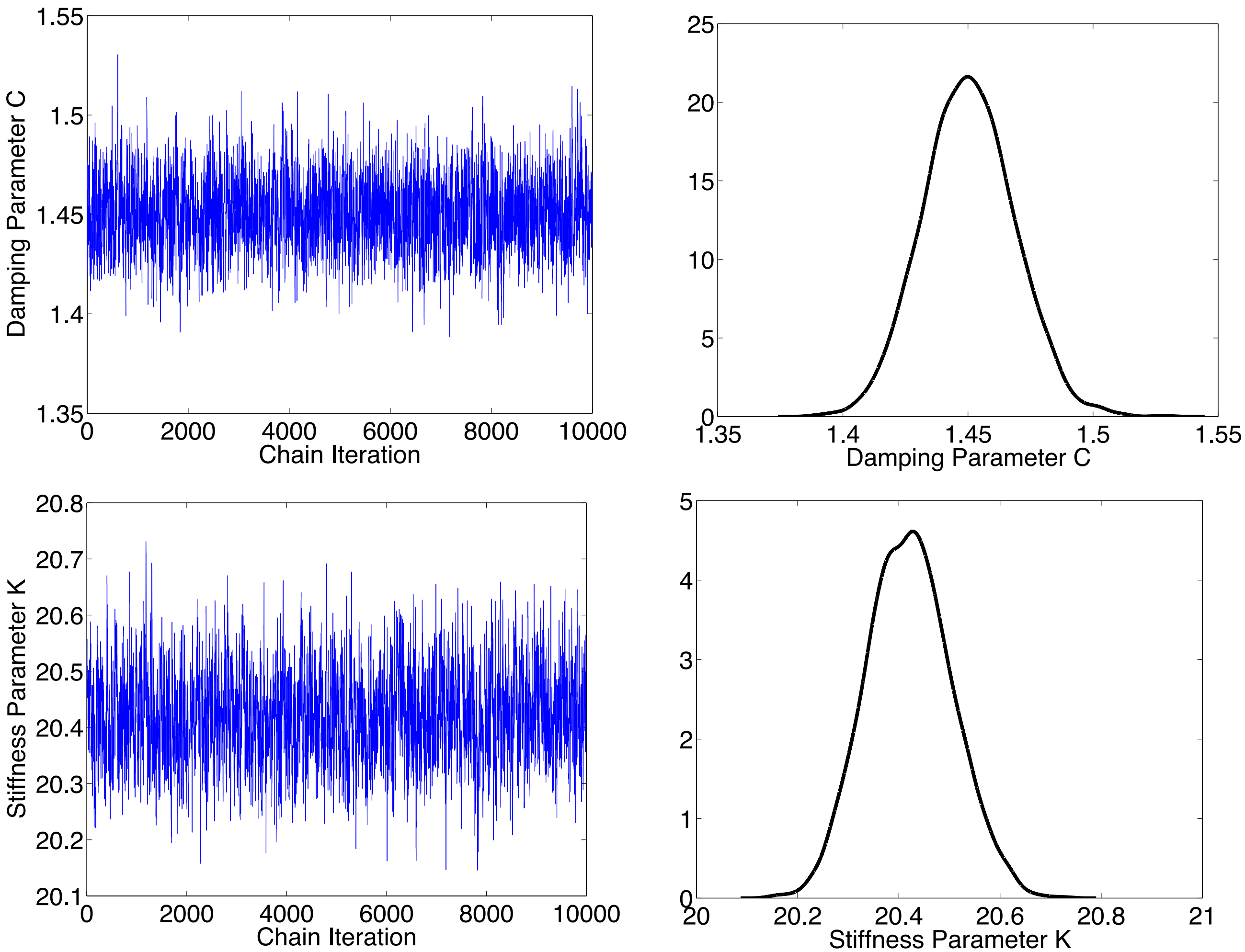


Note: Kernel density estimator (KDE) used to construct density.

Random Walk Metropolis

Case ii: Take $Q = [C, K, \sigma^2]$ with $J(q^*|q^{k-1}) = N(q^{k-1}, V)$ and

$$V = \begin{bmatrix} 0.000345 & 0.000268 \\ 0.000268 & 0.007071 \end{bmatrix}$$



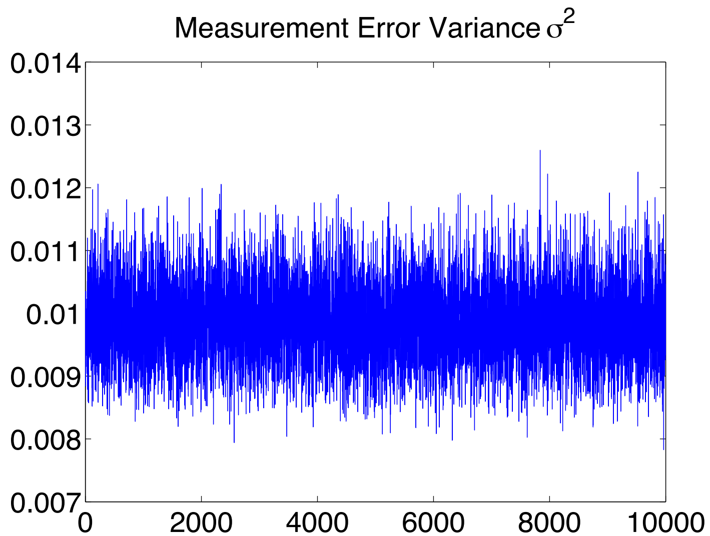
Note:

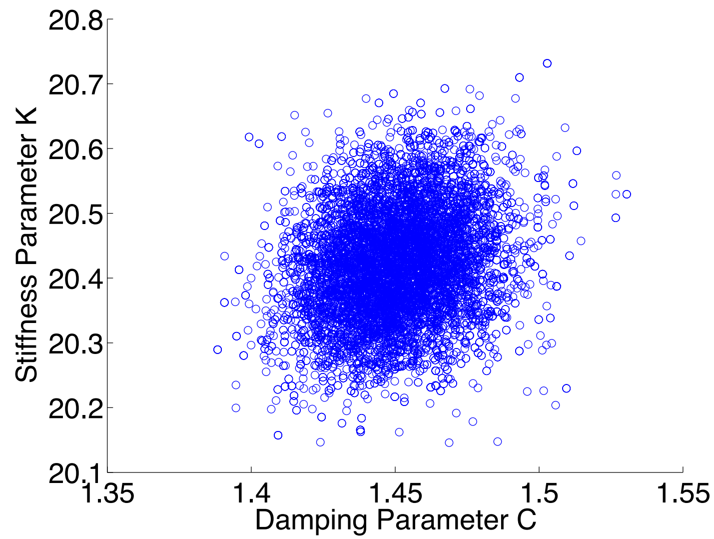
$$2\sigma_C \approx 0.04$$
$$\Rightarrow \sigma_C^2 \approx 0.4 \times 10^{-3}$$

$$2\sigma_K \approx 0.18$$
$$\Rightarrow \sigma_K^2 \approx 0.0081$$

Random Walk Metropolis

Case ii: Measurement error variance and joint samples



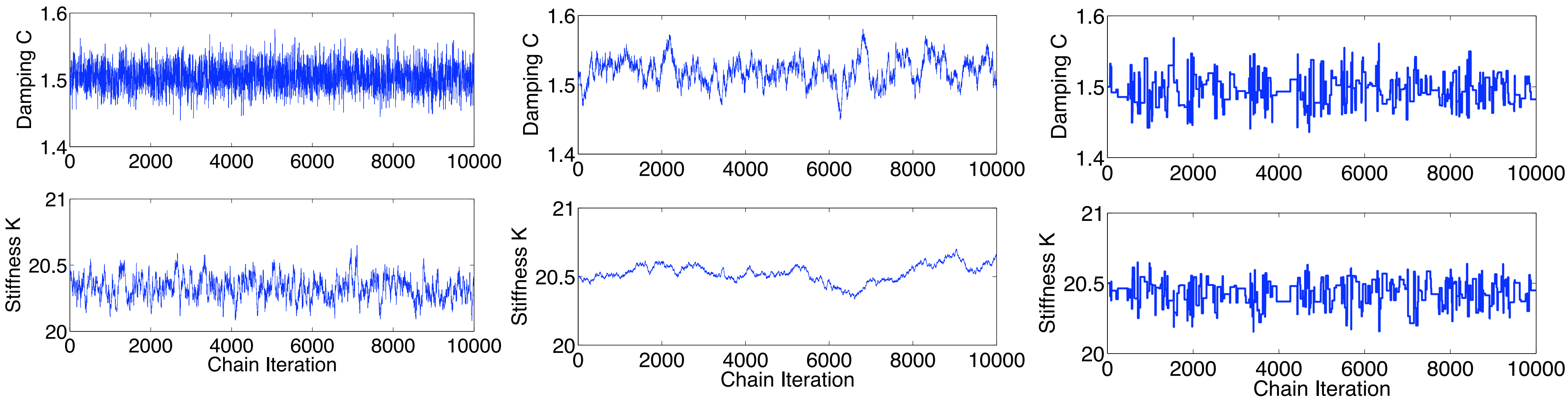


Codes:

- http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html
- spring_mcmc_C.m
- Spring_mcmc_C_K_sigma.m

Random Walk Metropolis

Case iii: Isotropic proposal function $J(q^*|q^{k-1}) = N(q^{k-1}, sI)$



$s = 9 \times 10^{-4}$ $\quad$ $s = 9 \times 10^{-6}$ $\quad$ $s = 9 \times 10^{-2}$

Stationary Distribution and Convergence Criteria

Here

$$
\begin{aligned}
p_{k-1,k} = \quad & P(X_k = q^k | X_{k-1} = q^{k-1}) \\
= \quad & P(\text{proposing } q^k) P(\text{accepting } q^k) \\
= \quad & J(q^k | q^{k-1}) \alpha(q^k | q^{k-1}) \\
= \quad & J(q^k | q^{k-1}) \min\left(1, \frac{\pi(q^k|\upsilon) J(q^{k-1}|q^k)}{\pi(q^{k-1}|\upsilon) J(q^k|q^{k-1})}\right)
\end{aligned}
$$

Detailed Balance Condition:

$$\pi_{k-1} p_{k-1,k} = \pi_k p_{k,k-1}$$

$$\Rightarrow \pi(q^{k-1}|\upsilon) p_{k-1,k} = \pi(q^k|\upsilon) p_{k,k-1}$$

From relation

$$\upsilon \min(1, x/\upsilon) = \min(x, \upsilon) = x \min(1, \upsilon/x)$$

it follows that

$$
\begin{aligned}
\pi(q^{k-1}|\upsilon) p_{k-1,k} = \quad & \pi(q^{k-1}|\upsilon) J(q^k|q^{k-1}) \min\left(1, \frac{\pi(q^k|\upsilon) J(q^{k-1}|q^k)}{\pi(q^{k-1}|\upsilon) J(q^k|q^{k-1})}\right) \\
= \quad & \pi(q^k|\upsilon) J(q^{k-1}|q^k) \min\left(1, \frac{\pi(q^{k-1}|\upsilon) J(q^k|q^{k-1})}{\pi(q^k|\upsilon) J(q^{k-1}|q^k)}\right) \\
= \quad & \pi(q^k|\upsilon) p_{k,k-1}
\end{aligned}
$$

Delayed Rejection Adaptive Metropolis (DRAM)

Adaptive Metropolis:

• Update chain covariance matrix as chain values are accepted.

$$V_k = s_p \mathrm{cov}(q^0, q^1, \cdots, q^{k-1}) + \varepsilon I_p$$

• *Diminishing adaptation* and *bounded convergenc*e required since no longer Markov chain.

• Employ recursive relations

$$\begin{aligned} \bar{q}^k &= \frac{1}{k+1} \sum_{i=0}^{k} q^i \\ &= \frac{k}{k+1} \cdot \frac{1}{k} \sum_{i=0}^{k-1} q^i + \frac{1}{k+1} q^k \\ &= \frac{k}{k+1} \bar{q}^{k-1} + \frac{1}{k+1} q^k \end{aligned}$$

$$V_{k+1} = \frac{k-1}{k} V_k + \frac{s_p}{k} \left[k\bar{q}^{k-1}(\bar{q}^{k-1})^T - (k+1)\bar{q}^k(\bar{q}^k)^T + q^k(q^k)^T + \varepsilon I_p \right]$$

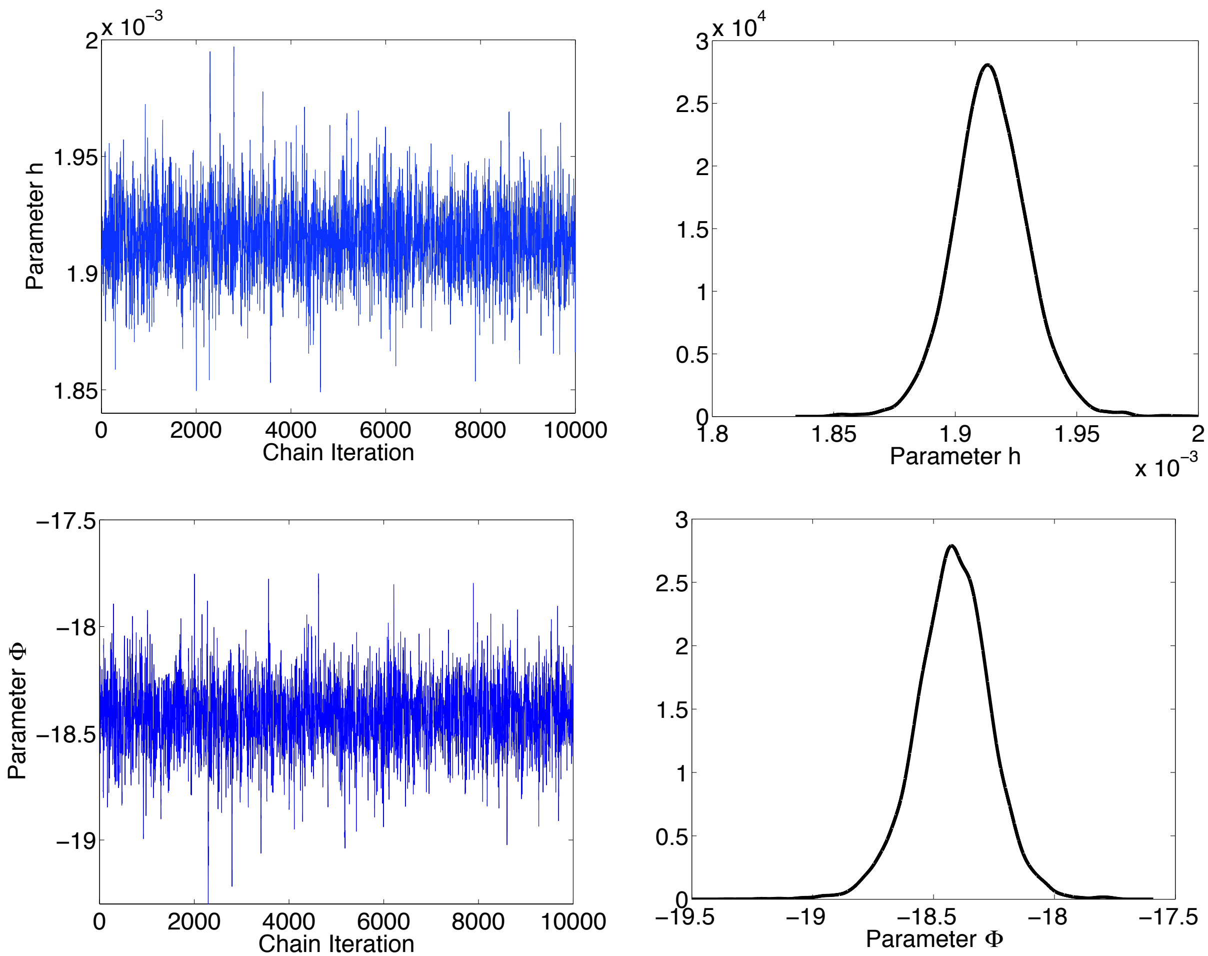
Delayed Rejection Adaptive Metropolis (DRAM)

Example: Heat model

$$\frac{d^2T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}\left[T_{amb} - T_s(L)\right]$$

Codes:
http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html



Bayesian Analysis

$$\sigma = 0.2604$$

$$\sigma_\Phi = 0.1552$$

$$\sigma_h = 1.5450 \times 10^{-5}$$

Frequentist Analysis

$$\sigma = 0.2504$$

$$\sigma_\Phi = 0.1450$$

$$\sigma_h = 1.4482 \times 10^{-5}$$

SIR Disease Example

SIR Model:

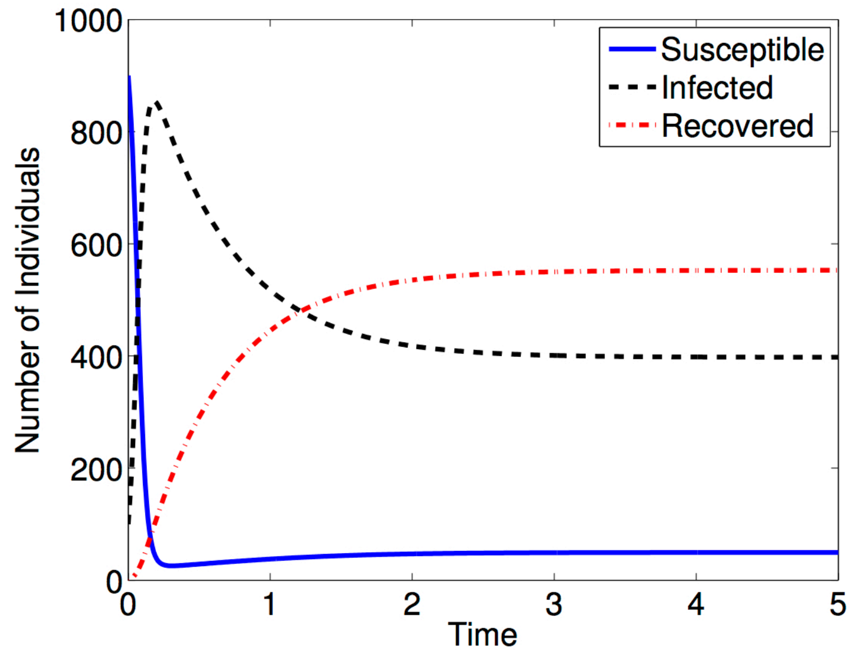
$$\frac{dS}{dt} = \delta N - \delta S - \underline{\gamma k} IS \quad , \quad S(0) = S_0 \qquad \text{Susceptible}$$

$$\frac{dI}{dt} = \underline{\gamma k} IS - (r + \delta) I \quad , \quad I(0) = I_0 \qquad \text{Infectious}$$

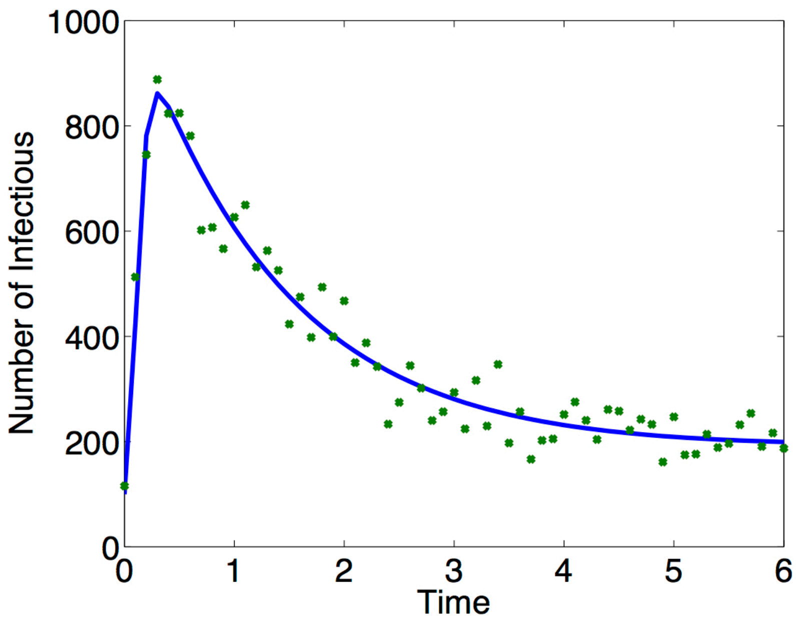
$$\frac{dR}{dt} = rI - \delta R \quad , \quad R(0) = R_0 \qquad \text{Recovered}$$

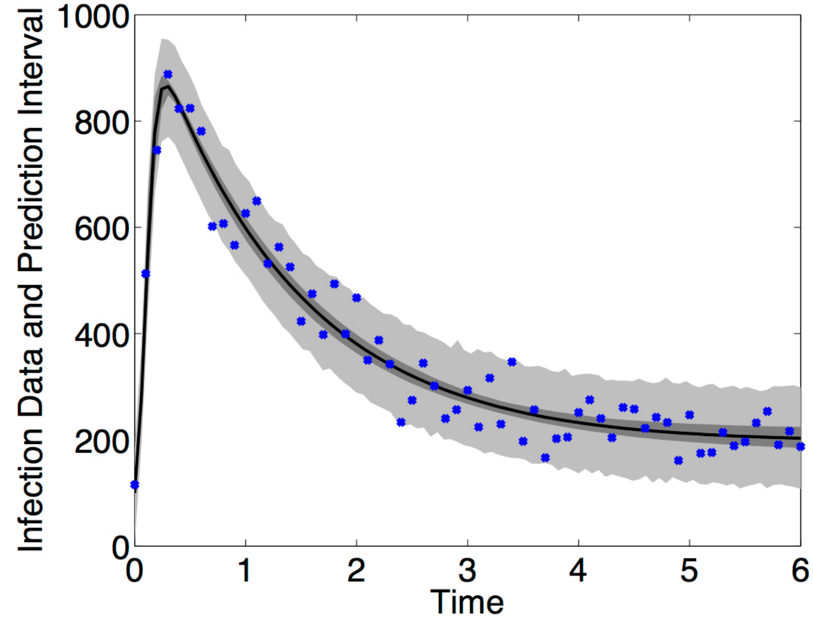
Note: Parameter set $q = [\gamma, k, r, \delta]$ is not identifiable

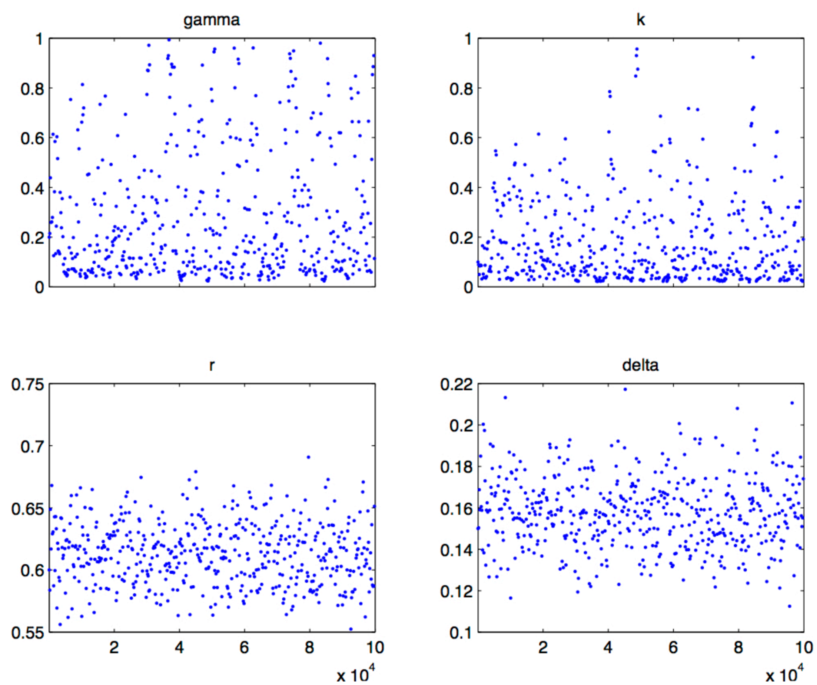
Typical Realization:

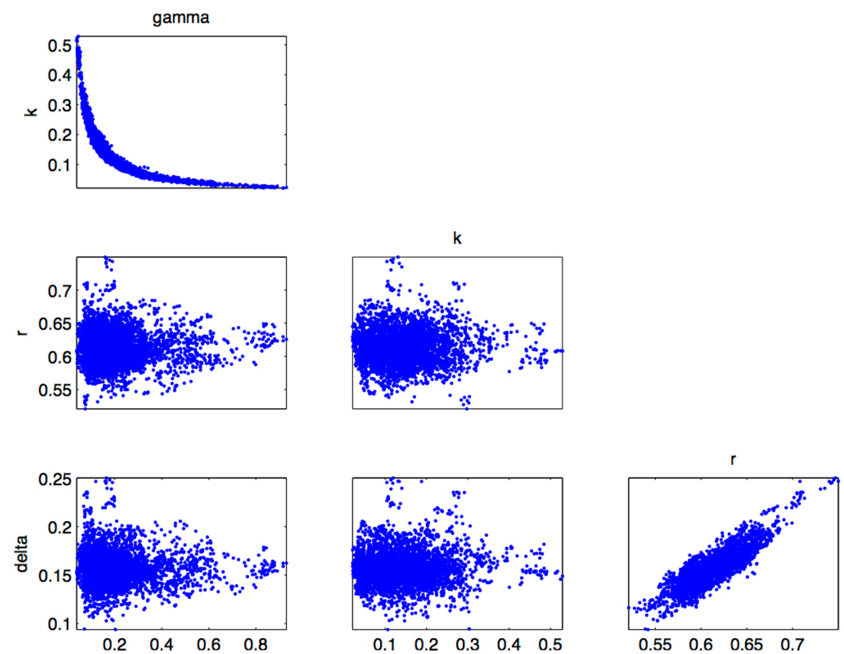


DRAM for SIR Example: Results









SIR Disease Example

Codes: 4 parameter case

- SIR_dram.m
- SIR_rhs.m
- SIR_fun.m
- SIRss.m
- mcmcpredplot_custom.m

Project problem: Modify for 3 parameter case

- SIR_dram.m
- SIR_rhs.m
- SIR_fun.m
- SIRss.m
- mcmcpredplot_custom.m

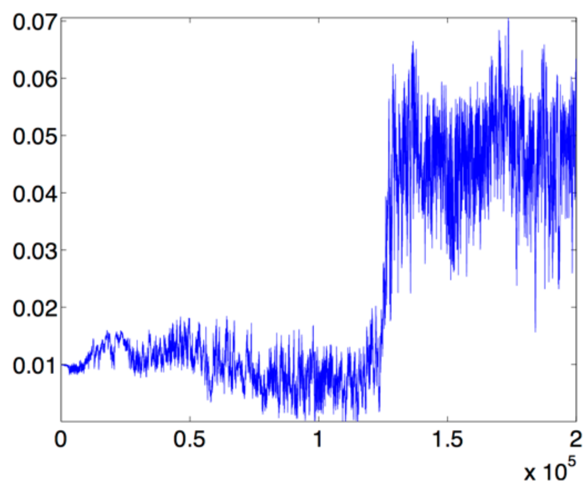
Bayesian Inference: Advantages and Disadvantages

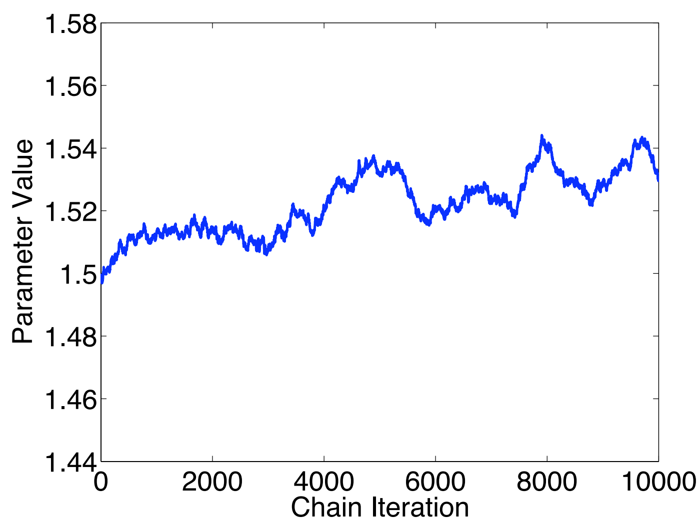
Advantages:

- Advantageous over frequentist inference when data is limited.
- Directly provides parameter densities, which can subsequently be propagated to construct response uncertainties.
- Can be used to infer non-identifiable parameters if priors are tight.
- Provides natural framework for experimental design.

Disadvantages:

- More computationally intense than frequentist inference.
- Can be difficult to confirm that chains have burned-in or converged.





Delayed Rejection Adaptive Metropolis (DRAM)

Websites

- http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html
- http://helios.fmi.fi/~lainema/mcmc/

Examples

- Examples on using the toolbox for some statistical problems.

Delayed Rejection Adaptive Metropolis (DRAM)

We fit the Monod model

$$y = \theta_1 \frac{1}{\theta_2 + 1} + \epsilon \quad , \quad \epsilon \sim N(0, I\sigma^2)$$

to observations

x (mg / L COD): 28 55 83 110 138 225 375

y (1 / h): 0.053 0.060 0.112 0.105 0.099 0.122 0.125

First clear some variables from possible previous runs.

```
clear data model options
```

Next, create a data structure for the observations and control variables. Typically one could make a structure data that contains fields xdata and ydata.

```
data.xdata = [28 55 83 110 138 225 375]'; % x (mg / L COD)
data.ydata = [0.053 0.060 0.112 0.105 0.099 0.122 0.125]'; % y (1 / h)
```

Construct model

```
modelfun = @(x,theta) theta(1)*x./(theta(2)+x);
ssfun = @(theta,data) sum((data.ydata-modelfun(data.xdata,theta)).^2);
model.ssfun = ssfun;
model.sigma2 = 0.01^2;
```

Delayed Rejection Adaptive Metropolis (DRAM)

Input parameters

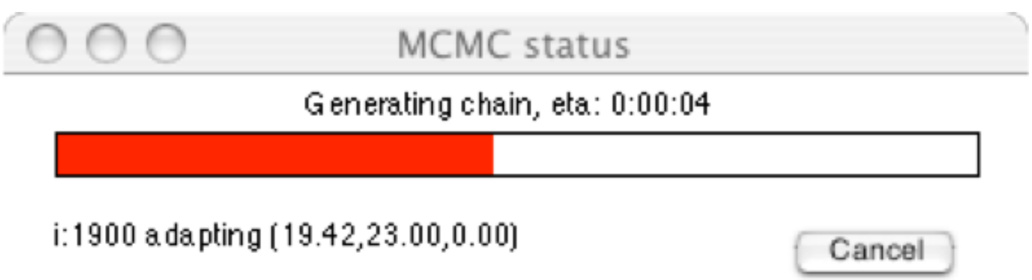
```
 params = {
      {'theta1', tmin(1), 0}
      {'theta2', tmin(2), 0} };
```

and set options

```
options.nsimu = 4000;
options.updatesigma = 1;
options.qcov = tcov;
```

Run code

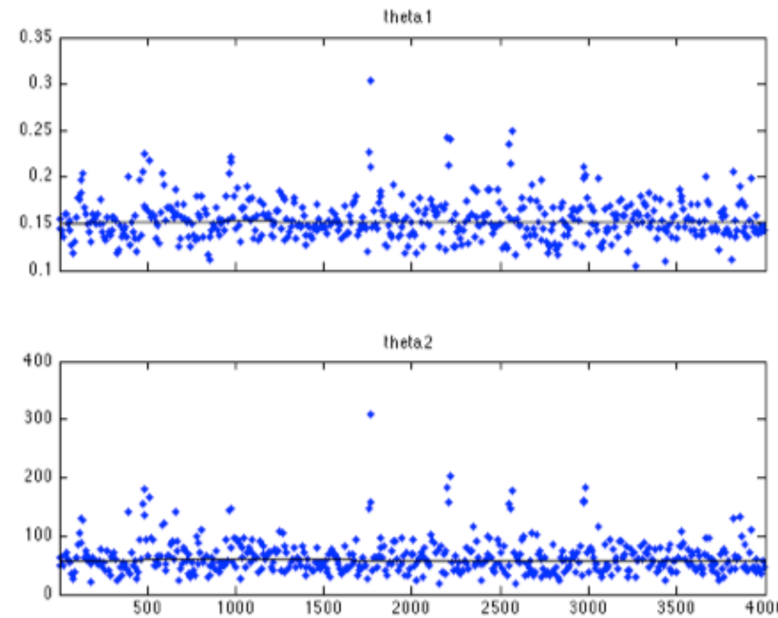
```
[res,chain,s2chain] = mcmcrun(model,data,params,options);
```

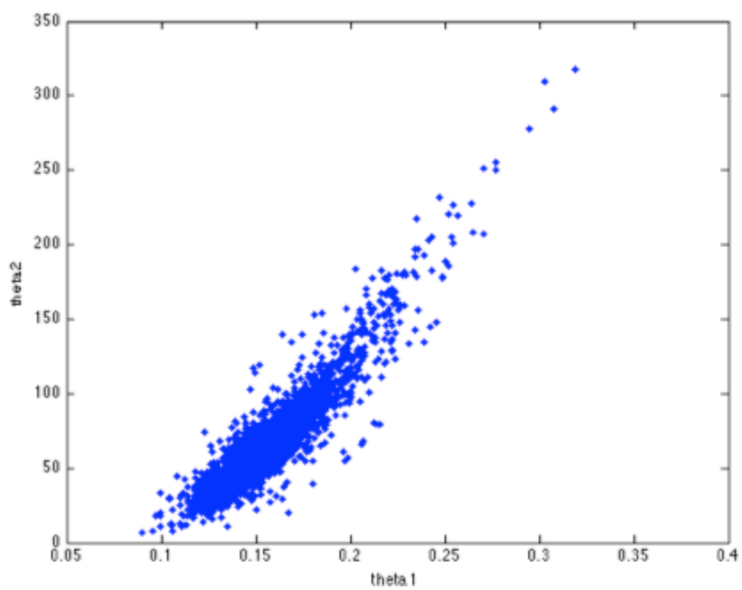


Delayed Rejection Adaptive Metropolis (DRAM)

Plot results

```
figure(2); clf
mcmcplot(chain,[],res,'chainpanel');
figure(3); clf
mcmcplot(chain,[],res,'pairs');
```





Examples:

•Several available in MCMC_EXAMPLES

•ODE solver illustrated in algae example

Delayed Rejection Adaptive Metropolis (DRAM)

Construct credible and prediction intervals

```
 figure(5); clf
 out = mcmcpred(res,chain,[],x,modelfun);
 mcmcpredplot(out);
 hold on
plot(data.xdata,data.ydata,'s'); % add data points to the plot
xlabel('x [mg/L COD]');
ylabel('y [1/h]');
hold off
title('Predictive envelopes of the model')
```