Bayesian Techniques for Parameter Estimation

"He has Van Gogh's ear for music," Billy Wilder

Reading: Sections 4.6 and 4.8, Chapter 8

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

Framework:

- Prior Distribution: Quantifies prior knowledge of parameter values.
- Likelihood: Probability of observing a data if we have a certain set of parameter values.
- Posterior Distribution: Conditional probability distribution of unknown parameters given observed data.

Joint PDF: Quantifies all combination of data and observations

$$\pi(q, \upsilon) = \pi(\upsilon|q)\pi_0(q)$$

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

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

Problem: Evaluation of normalization constant typically requires high dimensional integration.

Bayesian Inference

Uninformative Prior: No *a priori* information parameters

e.g., $\pi_0(q) = 1$ with limits

Informative Prior: Use conjugate priors; prior and posterior from same distribution

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

Evaluation Strategies:

- Analytic integration --- Rare
- Classical Gaussian quadrature; e.g., p = 1 - 4
- Sparse grid quadrature techniques; e.g., p = 5 - 40
- Monte Carlo quadrature Techniques
- Markov chain methods

Bayesian Inference

Example: Υ_i: Result from i^{th} coin toss

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

q: Probability of getting heads

Consider probability of observing series of tosses $\upsilon = [\upsilon_1, \cdots, \upsilon_N]$ given the probability q

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

N_1: Number of heads

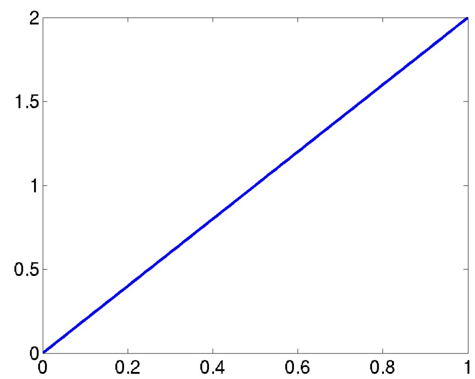
N_0: Number of tails

Uninformative prior yields

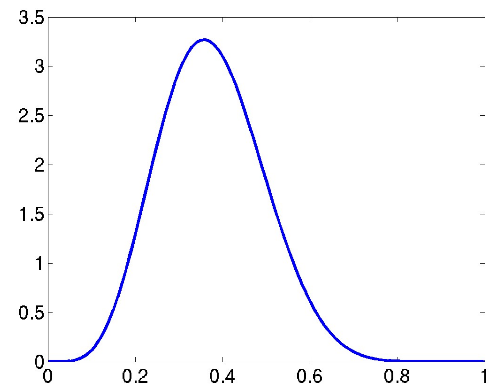
$$\pi(q|\upsilon) = \frac{q^{N_1}(1-q)^{N_0}}{\int_0^1 q^{N_1}(1-q)^{N_0} dq} = \frac{(N+1)!}{N_0! N_1!} q^{N_1}(1-q)^{N_0}.$$

Bayesian Inference

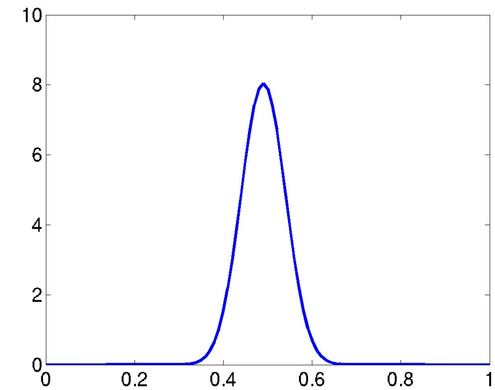
Example:



1 Head, 0 Tails



5 Heads, 9 Tails



49 Heads, 51 Tails

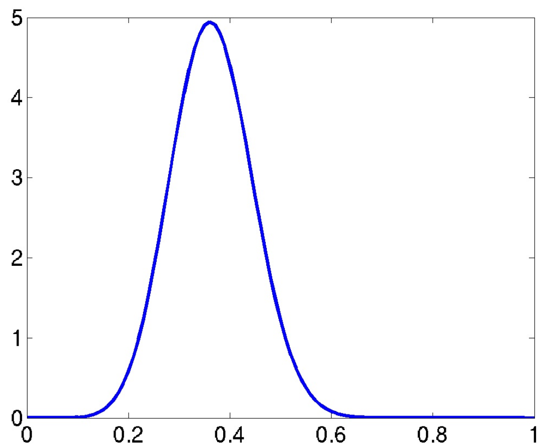
Note: For $N = 1$, frequentist theory would give probability 1 or 0

Bayesian Inference

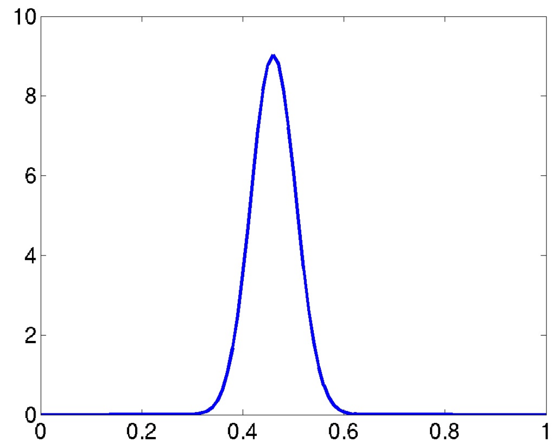
Example: Now consider

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

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

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

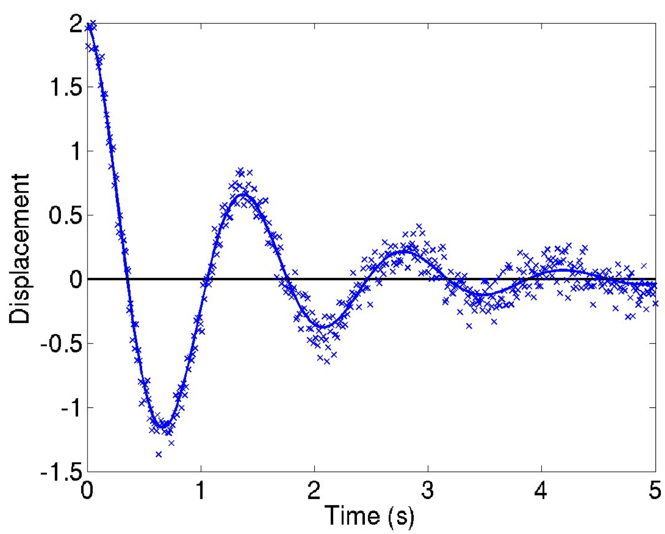
$$z(0) = 2\ ,\ \ \dot{z}(0) = -C$$

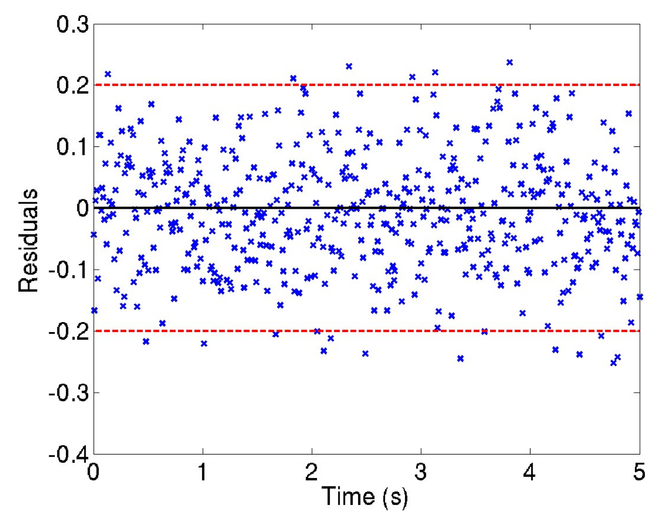
Note: Take $K = 20.5, C_0 = 1.5$

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

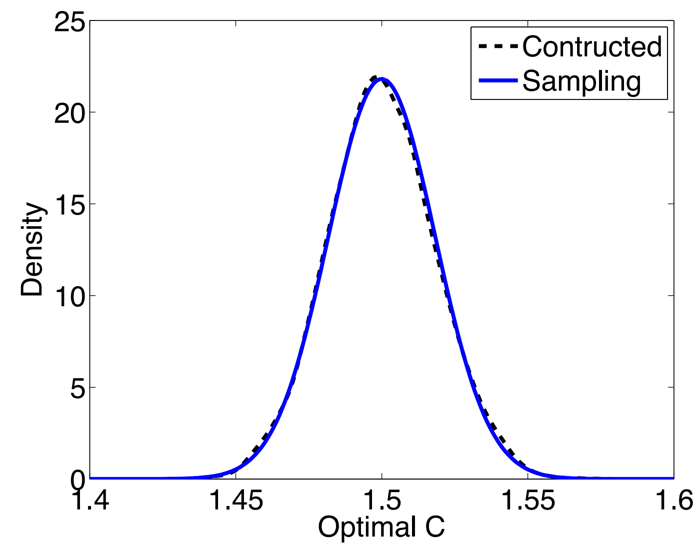
$$\frac{\partial y}{\partial C} = e^{-Ct/2}\left[\frac{Ct}{\sqrt{4K - C^2}}\sin\left(\sqrt{K - C^2/4}\cdot t\right) - t\cos\left(\sqrt{K - C^2/4}\cdot t\right)\right]$$

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

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

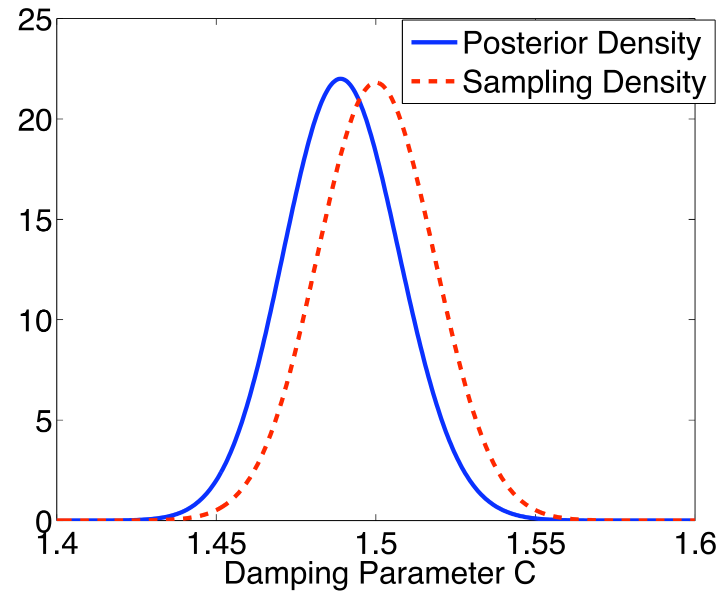
Posterior Distribution:

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

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.

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)

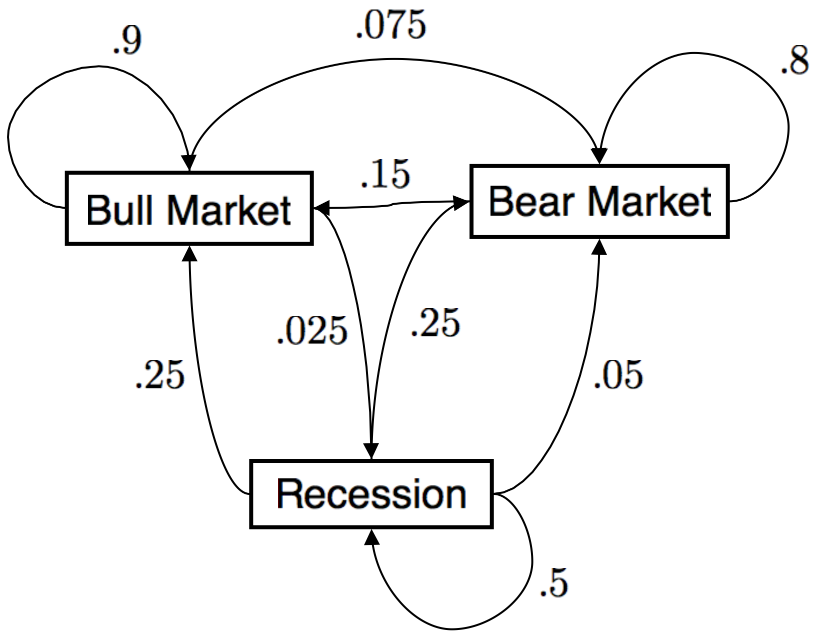
$$p_{ij} = P(X_{n+1} = x_j | X_n = x_i)$$

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

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

Markov Chains

Example:

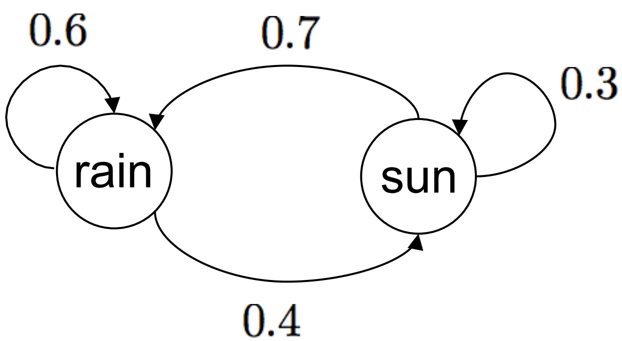


Chapman-Kolmogorov Equations: For any k such that $0 < k < n$,

$$p_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{(n-k)}$$

Markov Chains: Limiting Distribution

Example: Raleigh weather -- Tomorrow's weather conditioned on today's weather



$$P = \begin{bmatrix} .6 & .4 \\ .7 & .3 \end{bmatrix} \qquad S = \{\text{rain, sun}\}$$

- Distribution at Step n: $p^n = p^0 P^n$
- Note: Rows must sum to unity

Question:

(i) Can we say anything about $\lim_{n\to\infty} X_n$? Not really

(ii) What about $\lim_{n\to\infty} p^n = \pi$? Convergence in Distribution

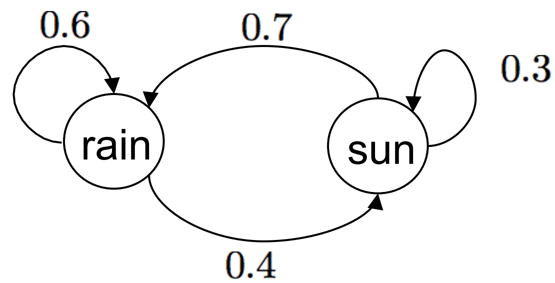
Note: If limit exists,

$$\pi = \lim_{n\to\infty} p^0 P^n = \lim_{n\to\infty} p^0 P^{n+1} = (\lim_{n\to\infty} p^0 P^n)P = \pi P$$

Definition: This is the limiting distribution (invariant measure)

Markov Chains: Limiting Distribution

Example: Raleigh weather



$$P = \begin{bmatrix} .6 & .4 \\ .7 & .3 \end{bmatrix}$$

Code: stationary.m

Solve

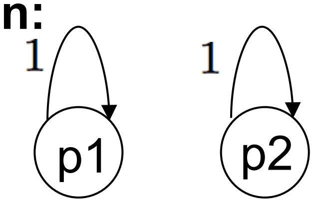
$$\pi P = \pi \quad , \quad \sum \pi_i = 1$$

$$\Rightarrow [\pi_r \ , \ \pi_s] \begin{bmatrix} .6 & .4 \\ .7 & .3 \end{bmatrix} = [\pi_r \ , \ \pi_s] \quad , \quad \pi_r + \pi_s = 1$$

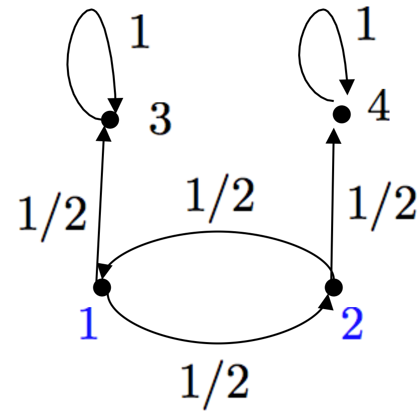
$$\pi_r = .6364 \ , \ \pi_s = .3636$$

Irreducible Markov Chains

Reducible Markov Chain:



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



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

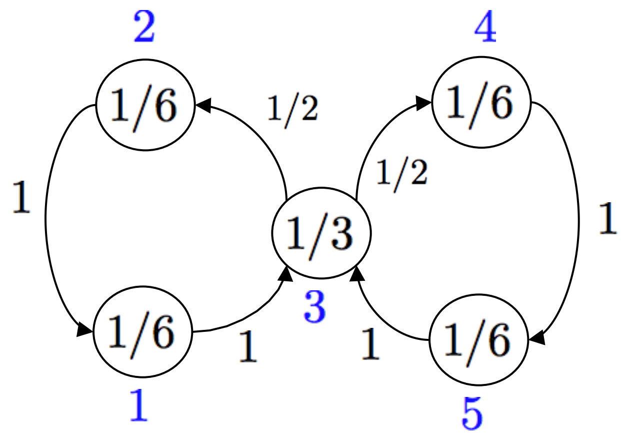
Note: Limiting distribution not unique if chain is reducible.

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

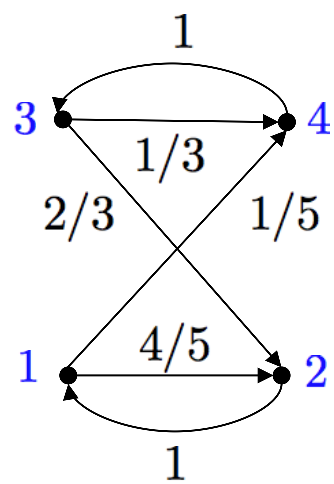
Periodicity: A Markov chain is *periodic* if parts of the state space are visited at regular intervals. The period *k* is defined as

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

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

Periodic Markov Chains

Example:



$$P = \begin{bmatrix} 0 & 4/5 & 0 & 1/5 \\ 1 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$p^0 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

$$p^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

Stationary Distribution

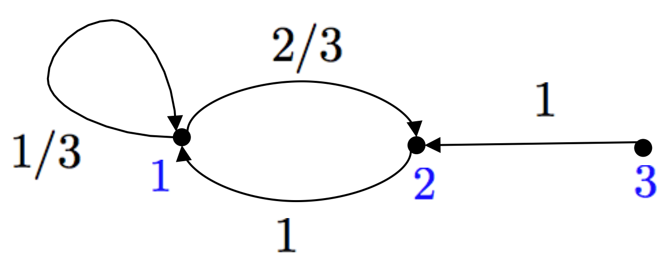
Theorem: A finite, homogeneous Markov chain that is irreducible and aperiodic has a unique stationary distribution π and the .chain will converge in the sense of distributions from any initial distribution p^0.

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

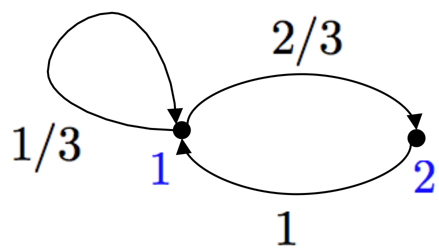
Definition: A matrix $A \in \mathbb{R}^{(n \times n)}$ is

(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

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

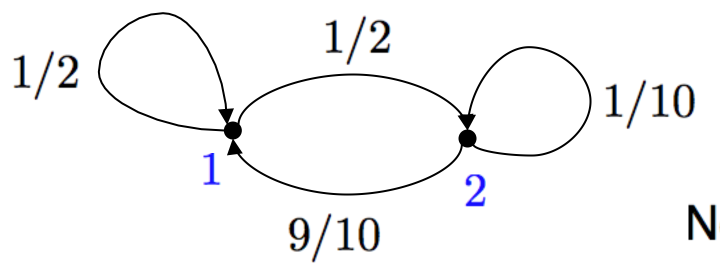
Detailed Balance Conditions

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

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

Proof: We need to show that $\pi_j = \sum_i \pi_i p_{ij}$. Note that $\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji}$

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.

Markov Chain Monte Carlo Methods

General Strategy:

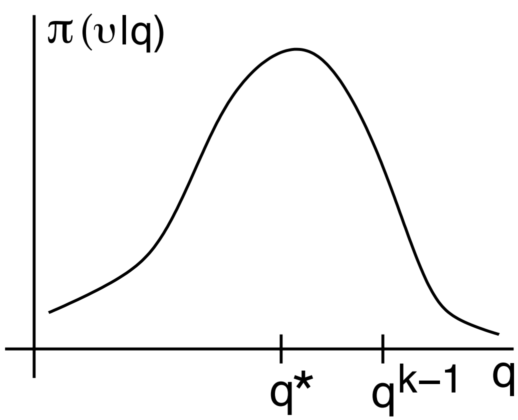
- Current value: $X_{k-1} = q^{k-1}$
- Propose candidate $q^* \sim J(q^*|q^{k-1})$ from proposal (jumping) distribution
- With probability $\alpha(q^*, q^{k-1})$, accept q^*; i.e., $X_k = q^*$
- Otherwise, stay where you are: $X_k = q^{k-1}$

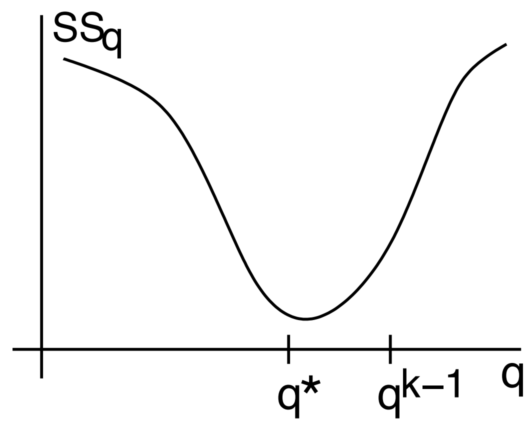
Intuition: Recall that

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

where

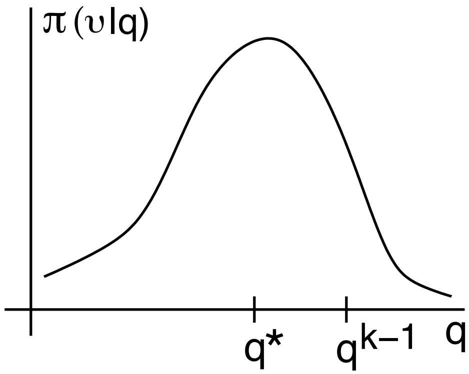
$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n [\upsilon_i - f_i(q)]^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

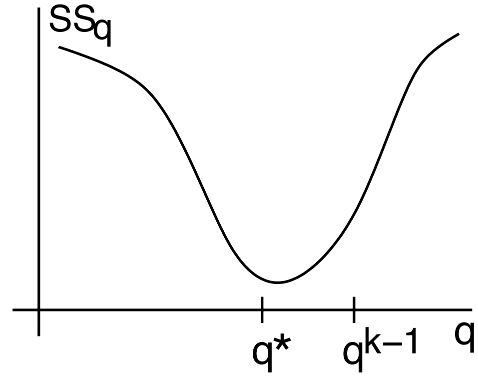




Markov Chain Monte Carlo Methods

Intuition:



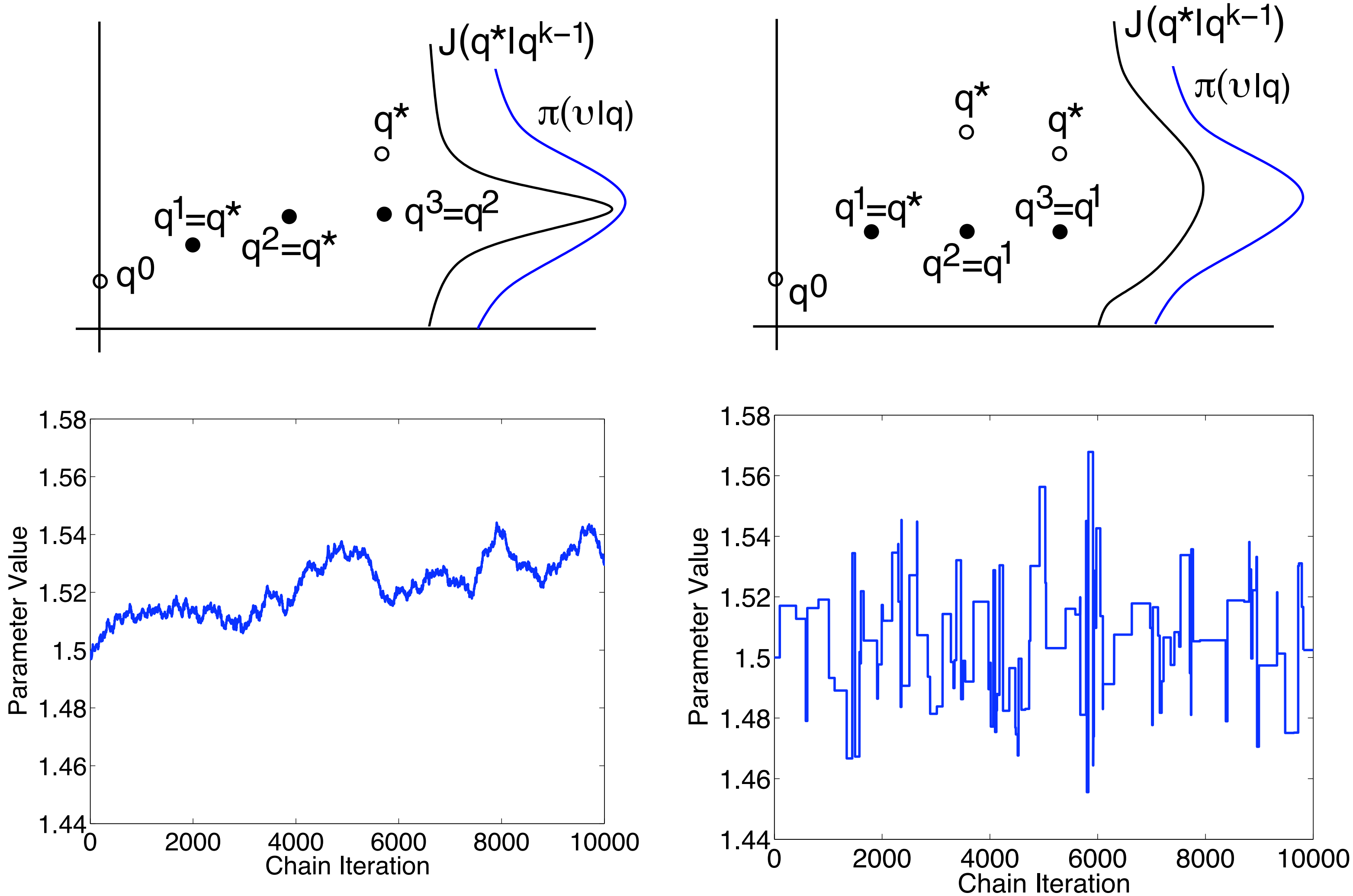


- Consider $r(q^*|q^{k-1}) = \frac{\pi(q^*|\upsilon)}{\pi(q^{k-1}|\upsilon)} = \frac{\pi(\upsilon|q^*)\pi_0(q^*)}{\pi(\upsilon|q^{k-1})\pi_0(q^{k-1})}$
  - If $r < 1 \Leftrightarrow \pi(\upsilon|q^*) < \pi(\upsilon|q^{k-1})$, accept with probability $\alpha = r$
  - If $r > 1$, accept with probability $\alpha = 1$

Note: Narrower proposal distribution yields higher probability of acceptance.

Markov Chain Monte Carlo Methods

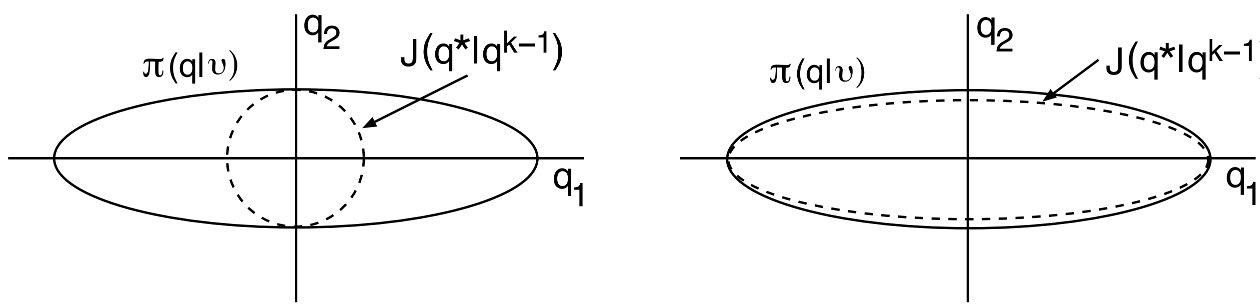
Note: Narrower proposal distribution yields higher probability of acceptance.



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: $z \sim N(0,1)$

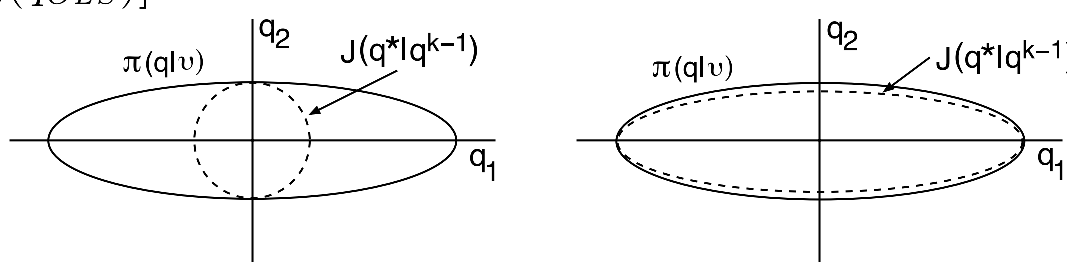
$$\text{(i) } R = cI \Rightarrow q^* \sim N(q^{k-1}, cI)$$

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

where

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

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



(a)
(b)

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

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

   (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 $r < 1$.

Metropolis-Hastings Algorithm

Metropolis-Hastings Algorithm: $J(q^*|q^{k-1})$ does not have to be symmetric

- Acceptance Ratio: $$r(q^*|q^{k-1}) = \frac{\pi(q^*|\upsilon)/J(q^*|q^{k-1})}{\pi(q^{k-1}|\upsilon)/J(q^{k-1}|q^*)}$$ $$= \frac{\pi(\upsilon|q^*)\pi_0(q^*)J(q^{k-1}|q^*)}{\pi(\upsilon|q^{k-1})\pi_0(q^{k-1})J(q^*|q^{k-1})}.$$

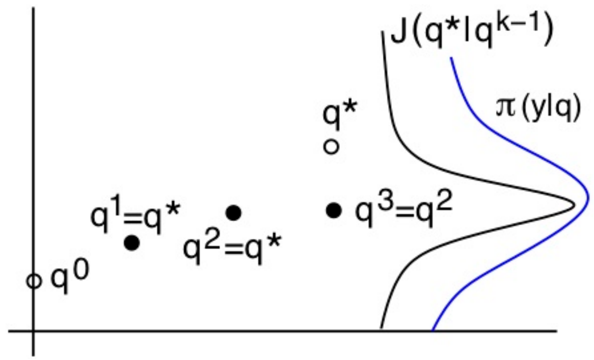
Examples:

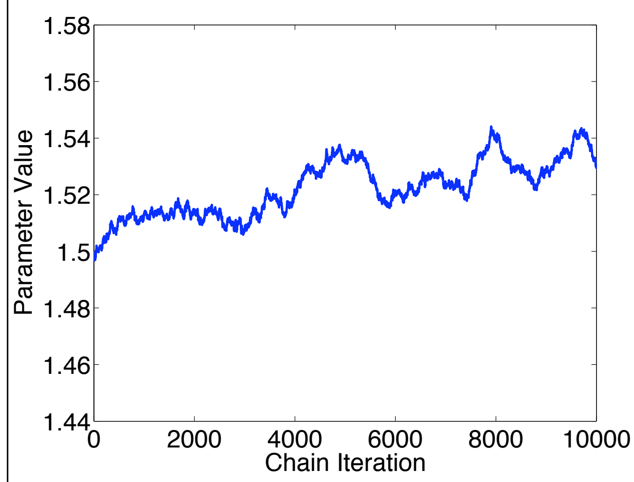
- Cauchy distribution: $J(q^*|q^{k-1}) = \frac{1}{\pi[1+(q^*)^2]}$
- $\chi^2(k)$ distribution: $J(q^*|q^{k-1}) = \kappa(q^*)^{k/2-1}e^{q^*/2}$

Note: Considered one of top 10 algorithms of 20th century

Random Walk Metropolis Algorithm for Parameter Estimation

1. Set number of chain elements M and design parameters n_s, σ_s
2. Determine $q^0 = \arg\min_q \sum_{i=1}^N [\upsilon_i - f_i(q)]^2$
3. Set $SS_{q^0} = \sum_{i=1}^N [\upsilon_i - f_i(q^0)]^2$
4. Compute initial variance estimate: $s_0^2 = \frac{SS_{q^0}}{n-p}$
5. Construct covariance estimate $V = s_0^2[\mathcal{X}^T(q^0)\mathcal{X}(q^0)]^{-1}$ and $R = \text{chol}(V)$
6. For $k = 1, \cdots, M$
   (a) Sample $z_k \sim N(0,1)$
   (b) Construct candidate $q^* = q^{k-1} + Rz_k$
   (c) Sample $u_\alpha \sim \mathcal{U}(0,1)$
   (d) Compute $SS_{q^*} = \sum_{i=1}^N [\upsilon_i - f_i(q^*)]^2$
   (e) Compute
   $$\alpha(q^*|q^{k-1}) = \min\left(1, e^{-[SS_{q^*} - SS_{q^{k-1}}]/2s_{k-1}^2}\right)$$
   (f) If $u_\alpha < \alpha$,
   Set $q^k = q^*$, $SS_{q^k} = SS_{q^*}$
   else
   Set $q^k = q^{k-1}$, $SS_{q^k} = SS_{q^{k-1}}$
   endif
   (g) Update $s_k \sim \text{Inv-gamma}(a_{val}, b_{val})$ where
   $$a_{val} = 0.5(n_s + n) \ , \ b_{val} = 0.5(n_s\sigma_s^2 + SS_{q^k})$$





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 $\sigma_s^2 = s_{k-1}^2 = \frac{R_{k-1}^T R_{k-1}}{n-p}$

Random Walk Metropolis

Example: We revisit the spring model

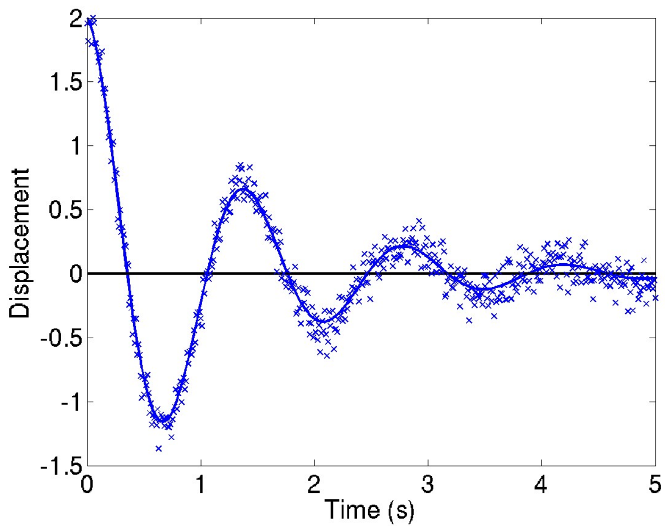
$$\ddot{z} + C\dot{z} + Kz = 0$$

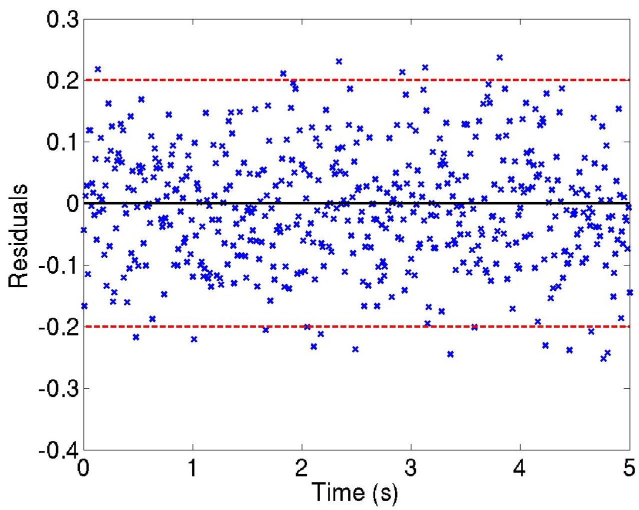
$$z(0) = 2 \; , \; \dot{z}(0) = -C$$

which has the solution

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

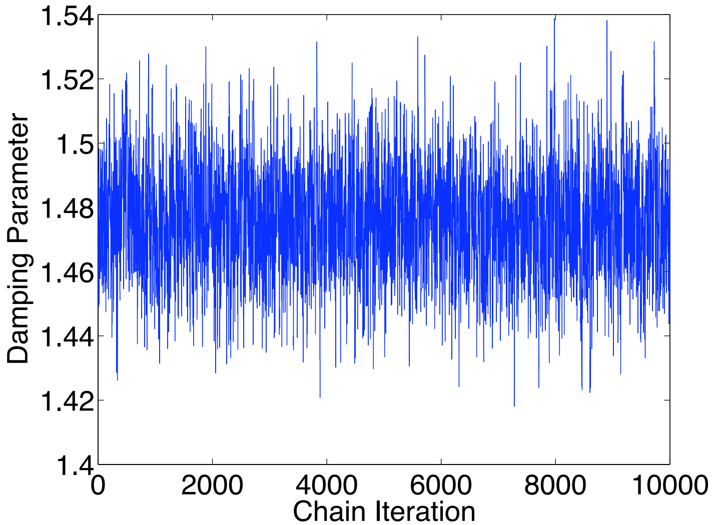
We assume that $\varepsilon_i \sim N(0, \sigma_0^2)$ where $\sigma_0 = 0.1$.

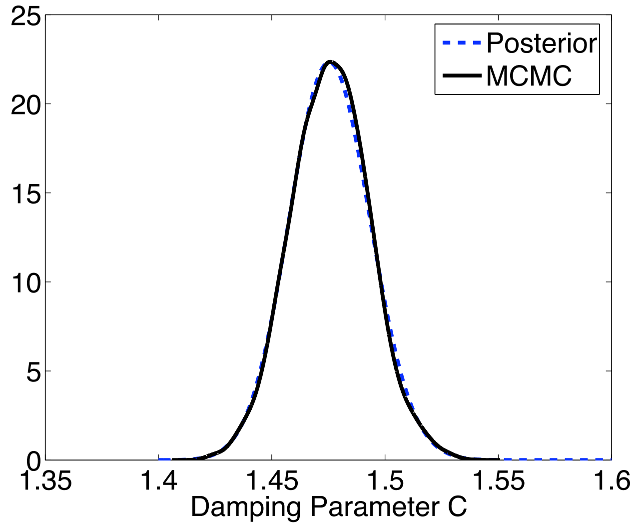




Random Walk Metropolis

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



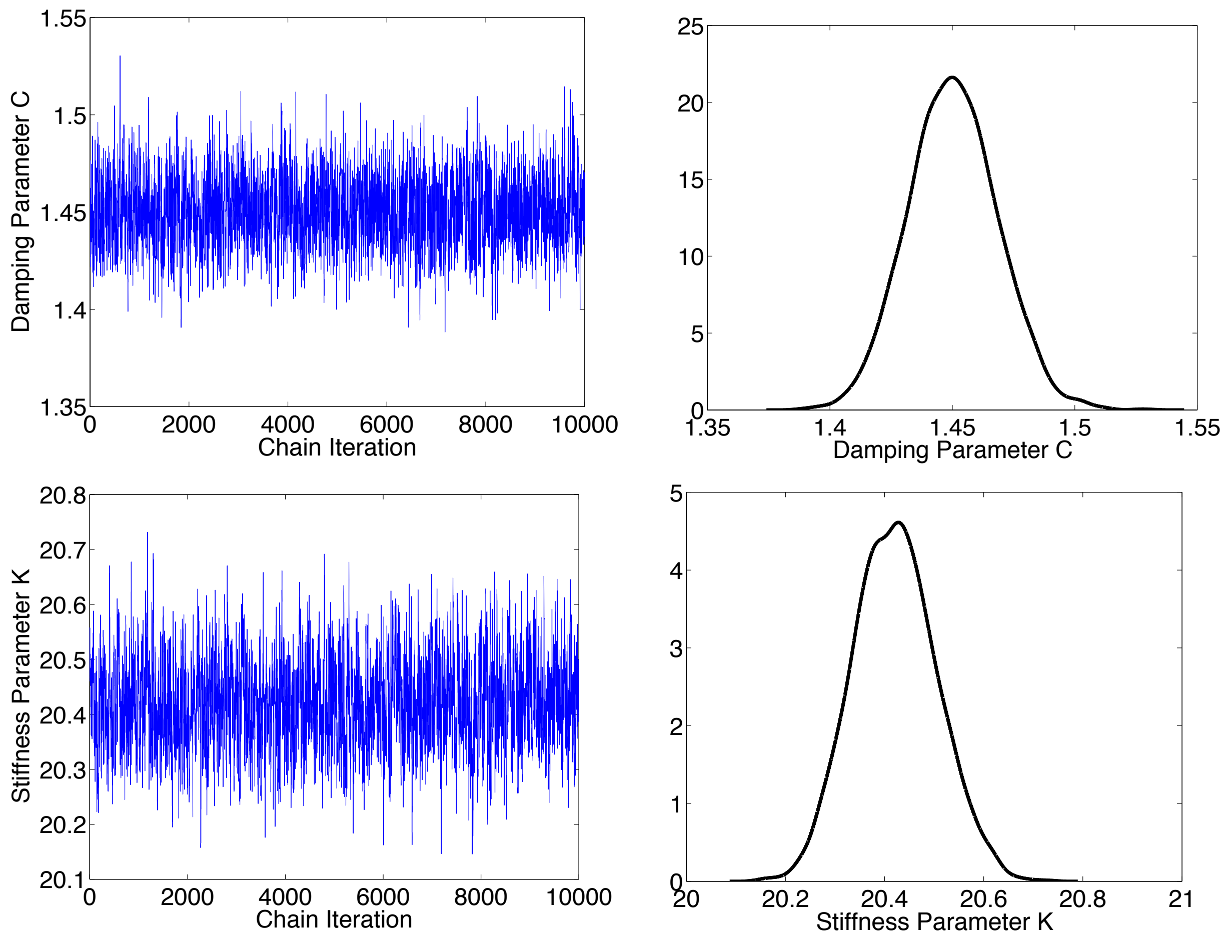


Note: Kernel density estimator (KDE) used to construct density.

Random Walk Metropolis

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

$$V = \left[\begin{array}{cc} 0.000345 & 0.000268 \\ 0.000268 & 0.007071 \end{array} \right]$$



Note:

$$2\sigma_C \approx 0.04$$

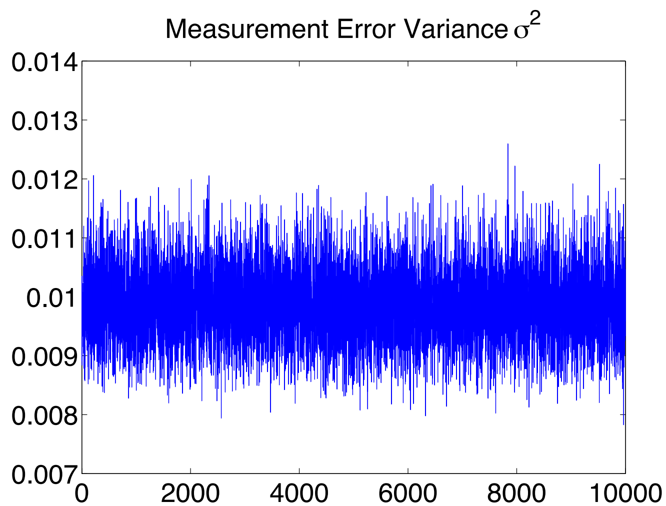
$$\Rightarrow \sigma_C^2 \approx 0.4 \times 10^{-3}$$

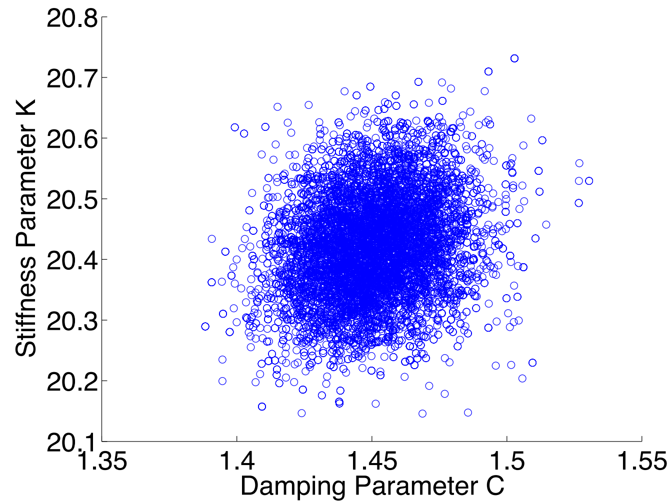
$$2\sigma_K \approx 0.18$$

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

Random Walk Metropolis

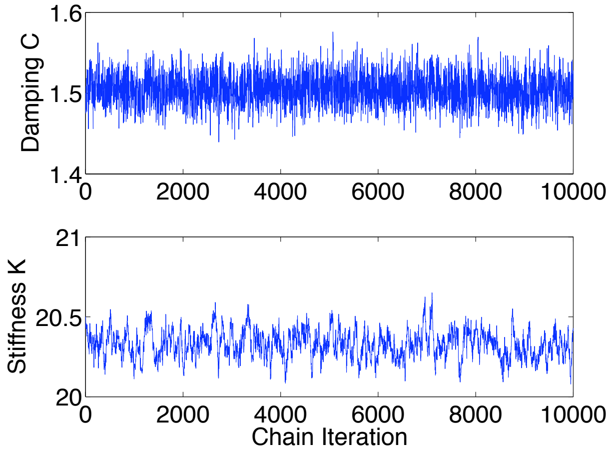
Case ii: Measurement error variance and joint samples



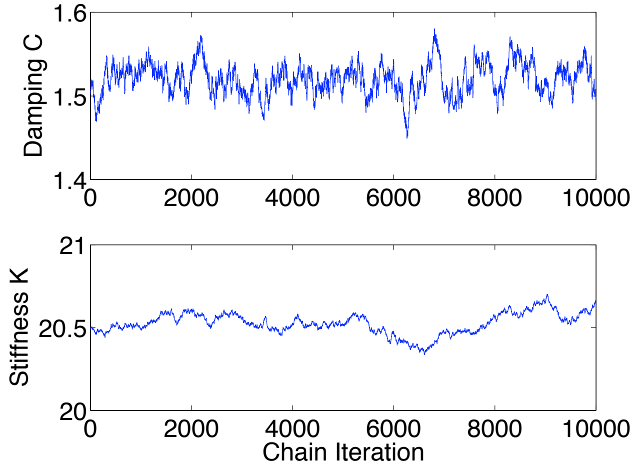


Random Walk Metropolis

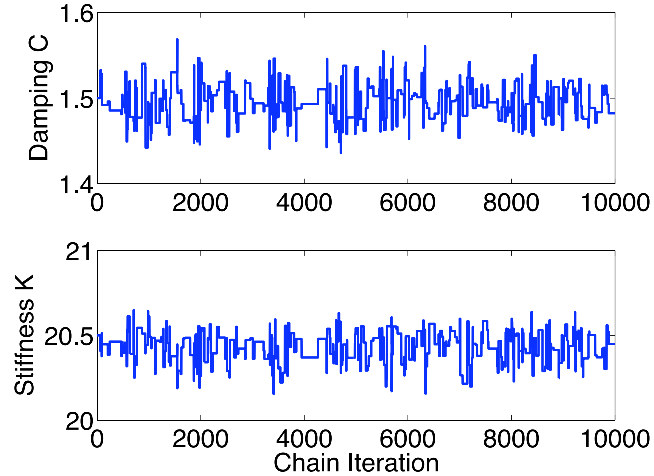
Case iii: Isotropic proposal function $J(q^*|q^{k-1}) = N(q^{k-1}, sI)$



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



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



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

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

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

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

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

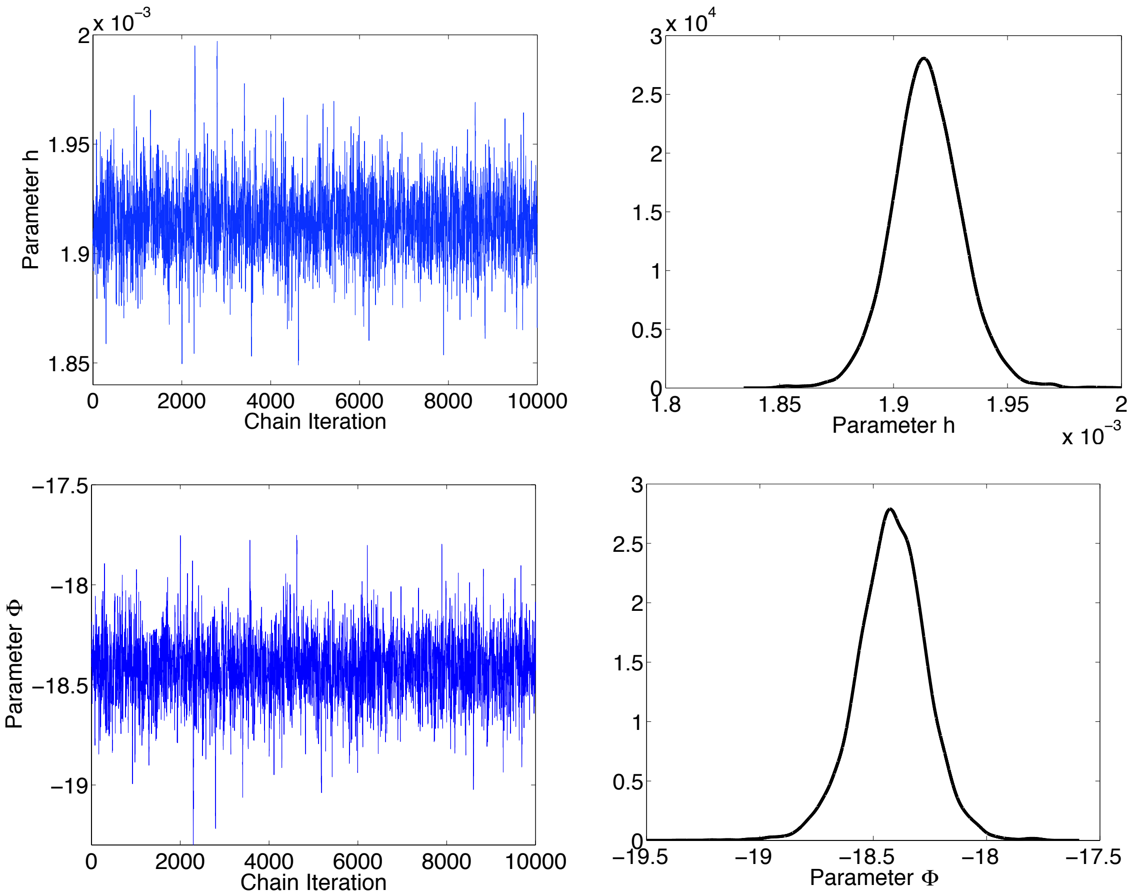
$$\begin{aligned}
\bar{q}^k = &\quad \frac{1}{k+1} \sum_{i=0}^{k} q^i \\
= &\quad \frac{k}{k+1} \cdot \frac{1}{k} \sum_{i=1}^{k-1} q^i + \frac{1}{k+1} q^k \\
= &\quad q^k + \frac{k}{k+1} \left(\bar{q}^{k-1} - q^k \right)
\end{aligned}$$

Delayed Rejection Adaptive Metropolis (DRAM)

Example: Heat model

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

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



Bayesian Analysis

$$\sigma = 0.2604$$

$$\sigma_\Phi = 0.1552$$

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

Frequentist Analysis

$$\sigma = 0.2504$$

$$\sigma_\Phi = 0.1450$$

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

Delayed Rejection Adaptive Metropolis (DRAM)

Example: HIV model

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

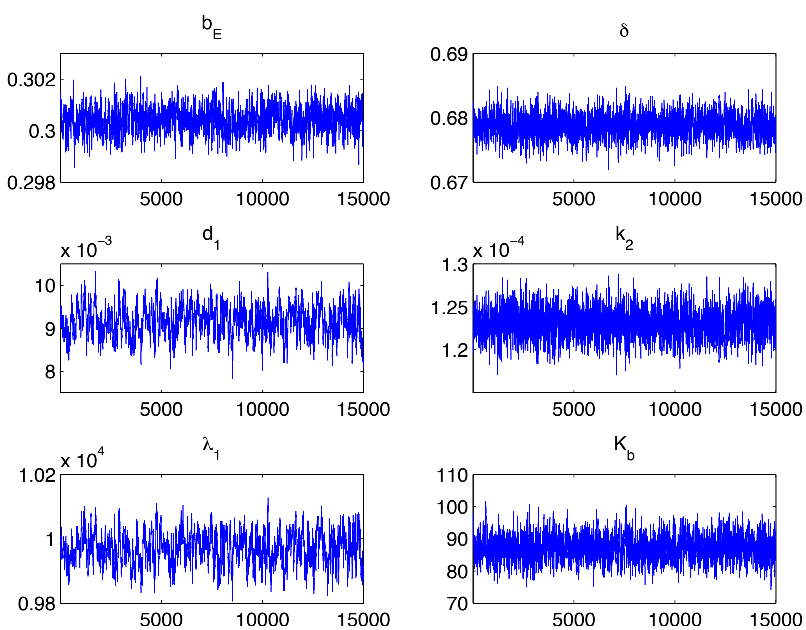
$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

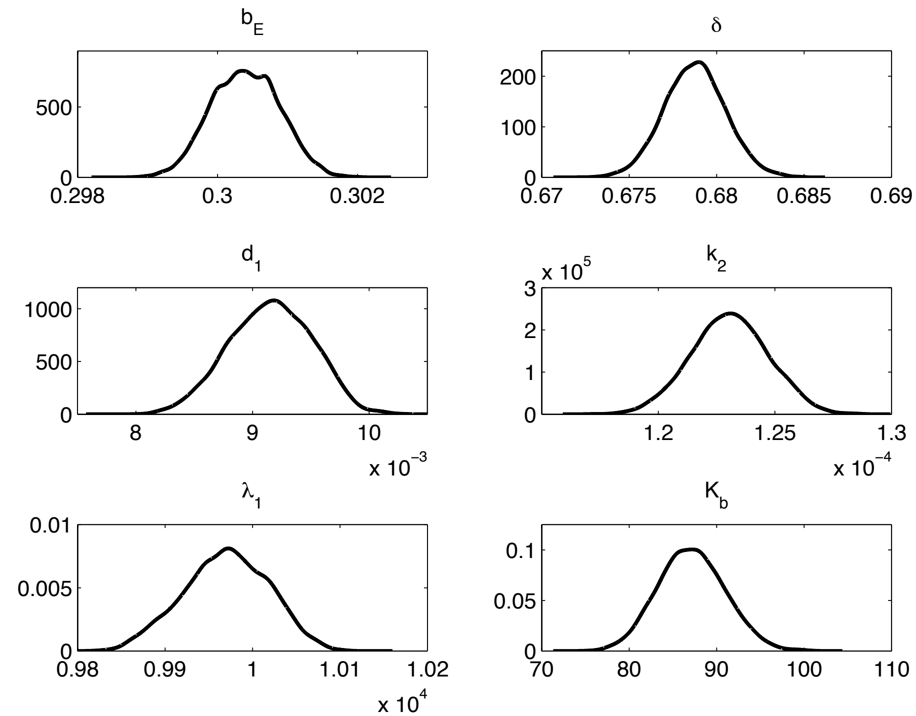
$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E.$$





Delayed Rejection Adaptive Metropolis (DRAM)

Example: HIV model

Note: Correlated versus nonidentifable parameters