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

\[ \pi(q, \nu) = \pi(\nu | q) \pi_0(q) \]

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

\[ \pi(q | \nu_{obs}) = \frac{\pi(\nu_{obs} | q) \pi_0(q)}{\pi(\nu_{obs})} = \frac{\pi(\nu_{obs} | q) \pi_0(q)}{\int_{\mathbb{R}^p} \pi(\nu_{obs} | q) \pi_0(q) dq} \]

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

**Uninformative Prior:** No *a priori* information parameters

\[ \pi_0(q) = 1 \text{ with limits} \]

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

\[ \pi(q|\text{obs}) = \frac{\pi(\text{obs}|q)\pi_0(q)}{\pi(\text{obs})} = \frac{\pi(\text{obs}|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\text{obs}|q)\pi_0(q) dq} \]

**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: $\gamma_i$: Result from $i^{th}$ coin toss

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

$q$: Probability of getting heads

Consider probability of observing series of tosses $v = [v_1, \cdots, v_N]$ given the probability $q$

$$\pi(v|q) = \prod_{i=1}^{N} q^{v_i} (1 - q)^{1-v_i}$$

$$= q^{\sum v_i} (1 - q)^{N - \sum v_i}$$

$$= q^{N_1} (1 - q)^{N_0}$$

$N_1$: Number of heads

$N_0$: Number of tails

Uninformative prior yields

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

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

with \( \mu = .3 \) and \( \sigma = .1 \).

**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:**

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

where

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

is the sum of squares error.
Parameter Estimation: Example

Example: Consider the spring model

\[ \ddot{z} + C \dot{z} + Kz = 0 \]

\[ z(0) = 2, \quad \dot{z}(0) = -C \]

which has the solution

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

Note: Take \( K = 20.5, C_0 = 1.5 \)

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

\[ \mathbf{\chi}(q) = \left[ \frac{\partial y}{\partial C}(t_1, q), \ldots, \frac{\partial y}{\partial C}(t_n, q) \right]^T \]

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

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

so that

\[ \hat{C} \sim N \left( C_0, \sigma_c^2 \right), \quad \sigma_c = 0.0183 \]
Parameter Estimation: Example

**Bayesian Inference:** Employ the uniformed prior

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

Posterior Distribution:

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

**Issue:** \(e^{-SS_q_{MAP}} \approx 3 \times 10^{-13}\)

**Midpoint formula:**

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

**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:

- Parameters considered to be random variables with associated densities.
  \[
  \pi(q|u) = \frac{\pi(u|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(u|q)\pi_0(q)\,dq}
  \]

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)

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

**Transition Probability:** (Markov Kernel)

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

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

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

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

Baseball: States are $S = \{\text{win}, \text{lose}\}$. 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.

![Transition Diagram]

- 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

\[
\pi = \pi P, \quad \sum \pi_i = 1
\]

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

to obtain

\[
\pi = [0.5714, 0.4286]
\]
Markov Chain Techniques

**Baseball Example:** Solve constrained relation

\[ \pi = \pi P, \quad \sum \pi_i = 1 \]

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

to obtain

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

**Alternative:** Iterate to compute solution

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p^n )</th>
<th>( n )</th>
<th>( p^n )</th>
<th>( n )</th>
<th>( p^n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[0.8000, 0.2000]</td>
<td>4</td>
<td>[0.5733, 0.4267]</td>
<td>8</td>
<td>[0.5714, 0.4286]</td>
</tr>
<tr>
<td>1</td>
<td>[0.6400, 0.3600]</td>
<td>5</td>
<td>[0.5720, 0.4280]</td>
<td>9</td>
<td>[0.5714, 0.4286]</td>
</tr>
<tr>
<td>2</td>
<td>[0.5920, 0.4080]</td>
<td>6</td>
<td>[0.5716, 0.4284]</td>
<td>10</td>
<td>[0.5714, 0.4286]</td>
</tr>
<tr>
<td>3</td>
<td>[0.5776, 0.4224]</td>
<td>7</td>
<td>[0.5715, 0.4285]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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:

\[ p^0 = [p_1, p_2] = \pi \]

Note: Limiting distribution not unique if chain is reducible.

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

\[ p_{ij}^{(n)} > 0 \text{ for all states in finite } n \]
Periodic Markov Chains

Example:

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

$$ k = \gcd \left\{ n | P_{ii}^{(n)} > 0 \right\} $$

$$ = \gcd \left\{ n | P(X_{m+n} = x_i | X_m = x_i) > 0 \right\} $$

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

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

Note: Probability mass “cycles” through chain so no convergence
Periodic Markov Chains

Example:

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

\[ p^0 = \begin{bmatrix}
1/4 & 1/4 & 1/4 & 1/4 \\
\end{bmatrix} \]

\[ p^0 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
\end{bmatrix} \]
Stationary Distribution

**Theorem:** A finite, homogeneous Markov chain that is irreducible and aperiodic has a unique stationary distribution \( \pi \) 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

![Diagram showing a Markov chain with states 1, 2, and 3, and transition probabilities labeled 1/3, 2/3, 1, and 1.]

**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:**

![Diagram of a Markov chain with states 1 and 2, and transition probabilities 1/3, 2/3, and 1/3, 2/3, 7/9, 2/9]  

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

$$P_2 = \begin{bmatrix} 7/9 & 2/9 \\ 1/3 & 2/3 \end{bmatrix}$$
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) $\lambda_0$ has geometric and algebraic multiplicity 1

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

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

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

Then

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

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

Proof: By Lemma and Corollary 1, $P$ has a largest eigenvalue $\lambda_0 = 1$.

Since multiplicity is 1, unique $\pi$ such that $\pi P = \pi$ and $\sum_i \pi_i = 1$.

Convergence: Express

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

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

Note:

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

Note:

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

and

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

Thus

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

\[ = \begin{bmatrix} p_1^0 & \cdots & p_n^0 \end{bmatrix} \begin{bmatrix} 1 & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ 1 & \cdots & \pi_n \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \cdots \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \end{bmatrix} \]

\[ = \begin{bmatrix} \pi_1 \cdots \pi_n \end{bmatrix} \]

\[ = \pi \]
Detailed Balance Conditions

**Reversible Chains**: A Markov chain determined by the transition matrix $P = [p_{ij}]$ is reversible if there is a distribution $\pi$ 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:

![Graph of a Markov chain with states 1 and 2 and transition probabilities labeled as 1/2, 1/2, 1/10, 9/10, and 2.]

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

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

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

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

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

**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 $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

Likelihood:

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

where

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

is the sum of squares error.
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
\[
\pi(q|v) = \frac{\pi(v|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(v|q)\pi_0(q) dq}
\]
where
\[
\pi(v|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n [v_i - f_i(q)]^2 / 2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q / 2\sigma^2}
\]
Markov Chain Monte Carlo Methods

**Intuition:**

![Graph of π(υ|q) and SSq](image)

- Consider $r(q^*|q^{k-1}) = \frac{\pi(q^*|v)}{\pi(q^{k-1}|v)} = \frac{\pi(v|q^*)\pi_0(q^*)}{\pi(v|q^{k-1})\pi_0(q^{k-1})}$
  
  - If $r < 1 \Leftrightarrow \pi(v|q^*) < \pi(v|q^{k-1})$, 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.

\[
\begin{align*}
\pi(q\mid v) & \quad \text{with} \quad J(q^*\mid q^{k-1}) \\
\end{align*}
\]

(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
  - Independent Metropolis
- Random walk (local Metropolis)
  \[ q^* = q^{k-1} + Rz \]

- Two (of several) choices: \( z \sim N(0, 1) \)
  1. \( R = cI \Rightarrow q^* \sim N(q^{k-1}, cI) \)
  2. \( R = \text{chol}(V) \Rightarrow q^* \sim N(q^{k-1}, V) \)

where

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

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

Sensitivity Matrix

\[
\begin{align*}
\pi(q|v) & \quad J(q^*|q^{k-1}) \\
\pi(q|v) & \quad J(q^*|q^{k-1})
\end{align*}
\]
Metropolis Algorithm

**Metropolis Algorithm:** [Metropolis and Ulam, 1949]

1. Initialization: Choose an initial parameter value $q^0$ that satisfies $\pi(q^0|v) > 0$.
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
   $$r(q^*|q^{k-1}) = \frac{\pi(q^*|v)}{\pi(q^{k-1}|v)} = \frac{\pi(v|q^*)\pi_0(q^*)}{\pi(v|q^{k-1})\pi_0(q^{k-1})}. \quad (1)$$
   
   (c) Set
   $$q^k = \begin{cases} 
   q^*, & \text{with probability } \alpha = \min(1, r) \\
   q^{k-1}, & \text{else.} 
   \end{cases}$$
   
   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:**
  
  $r(q^*|q^{k-1}) = \frac{\pi(q^*|v)J(q^*|q^{k-1})}{\pi(q^{k-1}|v)J(q^{k-1}|q^*)}$
  
  $= \frac{\pi(v|q^*)\pi_0(q^*)J(q^{k-1}|q^*)}{\pi(v|q^{k-1})\pi_0(q^{k-1})J(q^*|q^{k-1})}$.

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 $n_s, \sigma_s$

2. Determine $q^0 = \arg \min_q \sum_{i=1}^N [v_i - f_i(q)]^2$

3. Set $SS_{q^0} = \sum_{i=1}^N [v_i - f_i(q^0)]^2$

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 $u_\alpha \sim U(0, 1)$
   
   (d) Compute $SS_{q^*} = \sum_{i=1}^N [v_i - f_i(q^*)]^2$
   
   (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  
   \[ a_{val} = 0.5(n_s + n), \quad b_{val} = 0.5(n_s\sigma^2 + SS_{q^k}) \]
Sampling Error Variance

**Strategy:** Treat error variance $\sigma^2$ 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

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

has the conjugate prior

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

The posterior is

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

so that

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

or

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

**Note:**

- $n_0$ taken small;
  - e.g., $n_0 = 1$ or $n_0 = .01$
- Take $\sigma^2_s = s^2_{k-1} = \frac{R^T_{k-1}R_{k-1}}{n-p}$
Random Walk Metropolis

Example: We revisit the spring model

\[ \ddot{z} + C \dot{z} + K z = 0 \]

\[ z(0) = 2, \quad \dot{z}(0) = -C \]

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.
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

![Graph showing measurement error variance and joint samples for stiffness parameter K and damping parameter C.](image-url)
**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
\[ p_{k-1,k} = P(X_k = q^k | X_{k-1} = q^{k-1}) = P(\text{proposing } q^k) P(\text{accepting } q^k) = J(q^k | q^{k-1}) \alpha(q^k | q^{k-1}) = J(q^k | q^{k-1}) \min \left( 1, \frac{\pi(q^k | \nu) J(q^{k-1} | q^k)}{\pi(q^{k-1} | \nu) J(q^k | q^{k-1})} \right) \]

Detailed Balance Condition:
\[ \pi_{k-1} p_{k-1,k} = \pi_k p_{k,k-1} \]
\[ \Rightarrow \pi(q^{k-1} | \nu) p_{k-1,k} = \pi(q^k | \nu) p_{k,k-1} \]

From relation
\[ \nu \min(1, x/\nu) = \min(x, \nu) = x \min(1, \nu/x) \]

it follows that
\[ \pi(q^{k-1} | \nu) p_{k-1,k} = \pi(q^{k-1} | \nu) J(q^k | q^{k-1}) \min \left( 1, \frac{\pi(q^k | \nu) J(q^{k-1} | q^k)}{\pi(q^{k-1} | \nu) J(q^k | q^{k-1})} \right) = \pi(q^k | \nu) J(q^{k-1} | q^k) \min \left( 1, \frac{\pi(q^{k-1} | \nu) J(q^k | q^{k-1})}{\pi(q^k | \nu) J(q^{k-1} | q^k)} \right) = \pi(q^k | \nu) p_{k,k-1} \]
Delayed Rejection Adaptive Metropolis (DRAM)

**Adaptive Metropolis:**

- Update chain covariance matrix as chain values are accepted.

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

- *Diminishing adaptation* and *bounded convergence* required since no longer Markov chain.

- Employ recursive relations

\[
\begin{align*}
\bar{q}^k &= \frac{1}{k+1} \sum_{i=0}^{k} q^i \\
&= \frac{k}{k+1} \cdot \frac{1}{k} \sum_{i=0}^{k-1} q^i + \frac{1}{k+1} q^k \\
&= \frac{k}{k+1} \bar{q}^{k-1} + \frac{1}{k+1} q^k
\end{align*}
\]

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

**Example:** Heat model

\[
\frac{d^2 T_s}{dx^2} = \frac{2(a + b) h}{ab} \frac{1}{k} \left[ T_s(x) - T_{amb} \right]
\]

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

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

\[
\begin{align*}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - (1 - \varepsilon)k_1 V T_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1 - f \varepsilon)k_2 V T_2 \\
\dot{T}_1^* &= (1 - \varepsilon)k_1 V T_1 - \delta T_1^* - m_1 E T_1^* \\
\dot{T}_2^* &= (1 - f \varepsilon)k_2 V T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V} &= N_T \delta (T_1^* + T_2^*) - c V - [(1 - \varepsilon)\rho_1 k_1 T_1 + (1 - f \varepsilon)\rho_2 k_2 T_2] V \\
\dot{E} &= \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta E E.
\end{align*}
\]
Delayed Rejection Adaptive Metropolis (DRAM)

Example: HIV model

Note: Correlated versus nonidentifiable parameters
Chain Convergence (Burn-In)

**Techniques:**
- Visually check chains
- Statistical tests
- Often abused in the literature

Chain not converged

Chain for nonidentifiable parameter
Delayed Rejection Adaptive Metropolis (DRAM)

Websites

• [http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html](http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html)

Examples

• [Examples](http://helios.fmi.fi/~lainema/mcmc/) on using the toolbox for some statistical problems.
Delayed Rejection Adaptive Metropolis (DRAM)

We fit the Monod model

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

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

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

and set options

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

Run code

```plaintext
[res,chain,s2chain] = mcmcrun(model,data,params,options);
```
Delayed Rejection Adaptive Metropolis (DRAM)

Plot results

```matlab
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

```matlab
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')
```